A COGNITIVE PERSPECTIVE ON LEARNING, DECISION-MAKING, AND TECHNOLOGY EVALUATIONS IN ORGANISATIONS

A thesis presented for the degree of
Doctor of Philosophy of Imperial College London

By

Stefano Benigni

Supervised by
Prof. Paola Criscuolo, Prof. Markus Perkmann

Dissertation Committee
Prof. Daniel A. Levinthal, Prof. Maurizio Zollo

Management & Entrepreneurship Department
Imperial College Business School
Imperial College London

June 2023

ABSTRACT

This dissertation examines how firms' selection of technological and R&D opportunities shape the performance of their innovation efforts. Managers select R&D investments in complex and uncertain environments where it is difficult to learn from past decisions. I examine this challenge using empirical and agent-based modelling methods and by focusing on three interrelated aspects: managers' individual learning processes, the adaptation of mental representations in complex environments, and the role of distributed expertise in group evaluations. In the first chapter, I propose an alternative explanation to how managers learn from experience that does not involve feedback and that is thus applicable to contexts where learning from feedback is difficult. I test this novel learning mechanism, termed 'representation learning', by analysing a large proprietary dataset of patent evaluations and termination decisions made by managers at a Fortune 500 firm. The second chapter explores further implications for performance of representation learning by means of an agent-based model of representation and policy search in rugged landscapes. This study examines how different representation search strategies affect decision-makers' adaptation in complex environments. Finally, the third chapter explores the performance of group evaluation processes when evaluators differ in the depth and breadth of their knowledge of the technologies being evaluated. This research contributes to management literature by shedding light on the cognitive processes underlying learning and decision-making in uncertain and complex environments. These findings also have practical implications for strategy research and practice concerning the management of uncertain R&D and technology investments.

DECLARATION OF ORIGINALITY

I, Stefano Benigni, hereby confirm that I am the author of this dissertation.

The development of the first chapter was supported by my supervisors Paola Criscuolo and Markus Perkmann, who are both co-authors of the related paper.

COPYRIGHT DECLARATION

To my father, Fabio.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

INTRODUCTION

This dissertation examines how firms' selection of technological and R&D opportunities shape the performance of their innovation efforts. The challenge managers face when selecting R&D investments is determined by the interplay of two fundamental issues. On the one hand, the commercial or strategic value of R&D opportunities is uncertain and depends on multiple interdependent factors. On the other, the extended time lags between investments and commercialisation make it difficult to learn from past decisions. Hence, managers rely on combinations of portfolio approaches and group decision processes to select uncertain technologies and R&D opportunities.

I study this challenge by examining individual and group decision processes. The first two chapters of this dissertation focus on the individual cognition of managers and on how they learn in these contexts. The third chapter examines how the diverse expertise of evaluators contributes to the performance of group evaluation processes.

In Chapter 1, I build on the managerial cognition and cognitive science literatures to derive a learning mechanism that can explain learning in contexts where feedback is noisy or unobserved. I propose that the relationship between decision experience and performance is mediated by changes in the structure of managers' mental representations. I test my theoretical predictions with a large dataset of written patent evaluations made by managers at a Fortune 500 firm between 1995 and 2015.

Chapter 2 explores further implications for performance of my proposed learning mechanism with an agent-based model. This study shows how decision-makers learn by incrementally refining simple mental representations and balancing the trade-offs between broad and narrow representation search strategies.

Finally, Chapter 3 leverages the above empirical setting to study the performance of group evaluations. Specifically, I explore the boundary conditions of the assumption that greater diversity of expertise among evaluators ensures broader access to knowledge and improves the quality of group decisions.

In Chapter 1 - *"Representation Learning: How Individuals Learn when Feedback is Noisy or Unobserved"*, we examine the decision performance of managers who routinely evaluate R&D opportunities. Performance feedback is often noisy or delayed in these contexts due to large time lags between investments and commercialisation. We ask how learning from experience can be explained when feedback is unobserved or has low informative value.

To address this question, we develop a learning mechanism that explains how cognition can adapt to the environment regardless of the quality of feedback. While uncertainty is generally thought to impair learning by deteriorating the informative value of feedback, we argue that the persistent cognitive costs that individuals sustain when making repeated, uncertain decisions induce them to refine their understanding of the environment's causal structure – i.e., to refine the categories of their mental representations. This happens regardless of whether feedback is noisy or unobserved because, as research in the neuro-cognitive sciences suggests, mental representations adapt to minimise the cognitive costs of decisions. In turn, changes in representations result in improved decision performance, especially when individuals are initially inexperienced, and representations are underspecified. Performance improves when representations adapt because new conceptual distinctions are constrained by prior observations and background knowledge and are consistent with current representations.

We find support for this learning mechanism, which we refer to as representation learning, in the patent evaluation context. We measure individuals' mental representations by analysing evaluation statements that portfolio managers in a Fortune 500 high-tech firm were required to write for each patent of the firm's portfolio. We developed NLP tools to code the evaluation criteria used by the firm's managers – or patent engineers - in over 40,000 evaluation statements and derive a measure of their cognitive complexity based on these criteria. We interviewed the firm's patent engineers and analysed internal documentation to further validate the coded criteria. This measure allows us to observe changes in mental representation over time for each individual. Finally, we collected data on commercialisation outcomes for all the patents in the portfolio, which we can compare against patent engineers' forecasts to observe both errors of commission (type I) and errors of omission (type II) over time.

In line with our proposed theory, we find that patent engineers with less evaluation experience held simpler mental representations and used increasingly refined and complex representations as the number of evaluations increased. In turn, we find that patent engineers more accurately forecasted the future value of patents as they accumulated evaluation experience and that changes in their mental representations mediated these performance improvements.

These findings have implications for our understanding of experiential learning and the role of mental representation in guiding decisions under uncertainty. Contrary to reinforcement learning, representation learning can explain how decision-makers learn from experience regardless of whether performance feedback is observed. While research has assumed that noisy or delayed feedback impairs learning, this mechanism implies that conditions of absent or poor feedback can also foster learning by inducing the adaptation of mental representations.

Thus, this study also contributes to the literature on mental representation by showing that representations adapt towards greater complexity in contexts characterised by low-quality feedback. A central question in this literature pertains to the performance implications of different degrees of representational complexity in complex environments. This question has prompted a debate about whether simple rules and heuristics rather than more accurate and complex representations may offer superior performance. We inform this debate by emphasising the process by which decision-makers adapt to complex environments by developing greater complexity.

In Chapter 2 - *"Representation Search Breadth: An Agent-based Simulation of Representation Learning,* I further explore the performance implications of representation learning using an agent-based computational model. An implication of representation learning is that individuals face a trade-off between developing highly refined yet narrow representations on the one hand and refining their representations to a lower degree yet more broadly across several dimensions of representation on the other. For instance, in the patent evaluation context, managers may have acquired advanced knowledge of the technological aspects of patents but only a simplified understanding of competitive or legal issues. In contrast, other managers with the same level of experience may have acquired a less advanced yet more balanced understanding of technological, competitive, legal and other dimensions related to patent value. These different distributions of expertise are reflected in the structure of managers' mental representations and have performance implications. Managers who are "specialists" in some narrow aspects of a decision problem develop a complex understanding of one or few dimensions and have a superior capacity to discriminate and identify highly valuable solutions along those dimensions. On the contrary, "generalists" show a relatively more limited capacity to identify highly valuable solutions but can search more broadly across several dimensions.

A vast literature on the topic presents conflicting views on the implications of specialist and generalist knowledge for the identification of superior solutions and opportunities and for the outputs of inventive and creative work. The distinction is rooted in the above strategic trade-off that all decision-makers and creative workers face: either invest limited time entirely within a specific knowledge domain and become a specialist in that domain or invest it across several domains and become a generalist. Current research is divided on which strategy leads to superior performance, as proponents on each side have presented compelling arguments and evidence supporting their views.

I explore the performance implications of broad and narrow representation search strategies in an NK model of mental representation and search. The central elements of this model are two landscapes, namely the true environment landscape and a simplified representation of it that agents can iteratively refine. While the true landscape is fixed during each simulation, agents can search for increasingly accurate representations as they adapt to the environment. The main parameter of the simulations is the breadth of agents' representation search strategies. Agents can search narrowly and refine their representations across a few dimensions, or they can search broadly and refine representations across several dimensions.

I find that the optimal representation search strategy is contingent on the complexity of the environment and the level of noise that affects feedback signals. Specifically, contrary to previous research, intermediate levels of search breadth are associated with high performance only for low to moderate levels of complexity. Highly complex and noisy environments demand narrow search strategies, while broad search strategies are optimal at all levels of complexity when noise is low or absent.

The second main set of results explores the relationship between the breadth of search strategies and the optimal degree of representational complexity. In line with recent findings in this research stream, I find that, counterintuitively, less accurate representations can outperform more accurate ones – i.e., that the optimal degree of representational complexity does not necessarily match the true complexity of the environment. However, I show that less accurate representations can outperform more accurate ones only for broad rather than narrow representation search strategies.

These findings contribute to research on the trade-off between specialist and generalist knowledge and the role of representational complexity in guiding adaptation over complex environments.

In Chapter 3 - *"Group Evaluation Accuracy: The Role of Depth and Breadth of Expertise in the Selection of Technologies"*, I examine the contingencies relating evaluators' domain expertise to the accuracy of group evaluations of technology. With reference to the empirical setting context of Chapter 1, patent engineers were individually responsible for evaluating patents but could solicit other evaluators, including the firm's technology experts and patent inventors, to contribute to group evaluations whenever they deemed it necessary. Levering the expertise of multiple evaluators is costly but is expected to improve the accuracy of evaluations. Specifically, it is generally assumed that greater diversity of expertise among the evaluators contributing to group evaluations provides broader access to knowledge and information that improves the quality and accuracy of evaluations.

However, the boundary conditions of this assumption remain understudied. Specifically, this study examines the above central tenet of group evaluations that diversity of expertise provides a wealth of knowledge and perspectives that collectively improve the quality of evaluations.

The main argument I will present is twofold. First, while research characterises expertise diversity only in terms of differences between the distributions of expertise of evaluators – i.e., in terms of the group's *breadth diversity* – I argue that differences between their expertise in the technology being evaluated – i.e., the group's *depth diversity* – play at least an equally important role in determining evaluation accuracy. Second, the effect of depth diversity on group accuracy is contingent on depth diversity. High breadth diversity is detrimental to accuracy when depth diversity is high but beneficial to accuracy when depth diversity is low.

I find empirical support for my predictions. I extend the patent evaluation dataset analysed in Chapter 1 with over 5,000 email exchanges between patent engineers and other evaluators pertaining to group evaluations. I measure evaluators' distribution of expertise across technology areas as the stock of experience accumulated evaluating patents in each area. I find that evaluators with high levels of expertise in the focal technology were more likely to overestimate and less likely to underestimate the future value of patents. In turn, groups composed of evaluators who all had either high or low expertise in the focal technology – i.e., low depth diversity groups - were less likely to evaluate the value of patents accurately than groups comprising evaluators with both high and low levels of expertise – i.e. high depth diversity groups. Finally, while individual expertise breadth and group breadth diversity were not associated with evaluation accuracy in our setting, breadth diversity attenuated the positive relationship between depth diversity and accuracy as predicted. That is, the positive effect of high depth diversity on accuracy was lower when evaluators had expertise across different areas – i.e., when breadth diversity was high. However, high breadth diversity improved accuracy when depth diversity was low.

A key implication of these findings is that aggregating more and more diverse knowledge from multiple evaluators with unique expertise does not necessarily improve the accuracy of evaluations as generally assumed. High diversity of expertise on depth and breadth dimensions can be detrimental to group evaluation accuracy. Further, counterintuitively, counterintuitively, groups of evaluators who all specialise in the focal technology would benefit not from the contributions of other evaluators with expertise but from the contributions of evaluators with less or no expertise.

Collectively, these studies contribute to our understanding of the relationship between different dimensions of knowledge and expertise and the performance of decisions under uncertainty. The first main contribution of this dissertation is to propose a mechanism that explains the adaptation of decision-makers' mental representations induced by the accumulation of decision experience in uncertain environments. Relatedly, this work illustrates some of the performance implications of changes in representations and different strategies that decision-makers can adopt to refine their understanding of complex decision environments. Secondly, this work sheds light on issues related to the aggregation of decision-makers' knowledge and expertise for group evaluations under uncertainty.

CHAPTER 1

REPRESENTATION LEARNING: HOW INDIVIDUALS LEARN WHEN PERFORMANCE FEEDBACK IS NOISY OR UNOBSERVED

ABSTRACT

Reinforcement learning is viewed as the central mechanism in organizational learning literature. However, reinforcement learning cannot explain how accumulating decision experience can lead to valuable learning when performance feedback is highly noisy or unobserved. We build on managerial cognition literature to introduce the notion of representation learning, which provides an alternative explanation of learning when individuals make repeated decisions under causal ambiguity. We argue that persistent exposure to causal ambiguity may promote learning because it induces individuals to refine their mental representations of the environment, regardless of whether performance feedback is observed. This effect is more pronounced in conditions of higher causal ambiguity and when individuals do not already have knowledge of the most relevant dimensions of the environment's causal structure. We find support for our theory in the context of patent evaluation and termination decisions, made by 146 intellectual property experts in a high-tech Fortune-500 firm over 15 years. Our study demonstrates that learning from experience can occur even when performance feedback is not observed, and we specify representation learning as an alternative learning mechanism, distinct from reinforcement learning. Our insights extend work on deliberate and mindful learning and contribute to research on the performance implications of complex representations.

INTRODUCTION

Theories of experiential learning, which explain how the accumulation of experience shapes current cognition and behaviour, are generally based on the idea of reinforcement learning (Levitt and March, 1988; Levinthal and Rerup, 2006; Argote and Miron-Spektor, 2011). The notion that individuals and organisations iteratively adjust to performance feedback has been widely adopted in the literature (Cyert and March, 1963; Nelson and Winter 1982; Greve, 2003). Nonetheless, recent work has challenged the assumption that accumulating experience generates feedback that systematically leads to valuable learning (Nelson, 2008; March, 2010). Experience, defined as the accumulation of task performances (Argote and Miron-Spektor, 2011), does not necessarily furnish the unambiguous and timely performance feedback that the reinforcement mechanism requires. Feedback information may be delayed (Denrell, Fang, and Levinthal, 2004; Rahmandad 2008) or difficult to interpret and impute to prior actions (Zollo, 2009; Levinthal and Rerup, 2021). Performance feedback may even be unobserved by decision makers; for instance, when delays are excessively large or due to information asymmetries (Mosakowski, 1997). This research maintains that in these situations of noisy or unobserved feedback, experiential learning, as explained by reinforcement, can lead to spurious associations between actions and outcomes (Denrell, 2008; Zollo, 2009).

The literature has examined alternatives to reinforcement learning, notably deliberate (Zollo and Winter, 2002) and mindful (Weick, Sutcliffe, and Obstfeld 1999; Levinthal and Rerup, 2006) learning. These theories challenge the automaticity associated with reinforcement learning. They argue that individuals and organizations decide to invest attention and resources in making sense of ambiguous information and improving their understanding of the causal structure of the environment. However, these prior contributions do not offer a complete explanation of experiential learning when feedback is noisy or unobserved, for two reasons. First, the accumulation of experience is not the main explanatory variable in these theories; and second, these theories still implicitly or explicitly explain learning in terms of performance feedback.

In this paper, we extend recent thinking on deliberate and mindful learning and propose a new mechanism called representation learning, which can explain experiential learning when performance feedback is highly noisy or unobserved. We focus on individuals who make repeated decisions under causal ambiguity (Mosakowski, 1997). Building on managerial cognition literature (Walsh 1995; Gavetti, 2005; Eggers and Kaplan, 2013) and cognitive science (Rosch et al., 1976; Radulescu, Shin, and Niv 2021), we propose that learning occurs in these contexts by means of changes in the representations that decision makers use to understand the causal structure of the environment (Fiske and Taylor, 1984; Barr, Stimpert, and Huff, 1992; Martignoni, Menon, and Siggelkow, 2016; Csaszar and Ostler, 2020). We show that by repeatedly making decisions – and being exposed to the causal ambiguity that decisions entail – individuals learn because they are induced to increase the accuracy of their mental representations. Contrary to reinforcement learning, these changes in cognition are prompted by repeated perceptions of causal ambiguity and occur regardless of whether performance feedback is observed.

Our argument comes in two steps. First, repeated exposure to the cognitive efforts that causal ambiguity demands, induces decision makers to refine their mental representations of the environment by drawing new conceptual distinctions; this occurs because the degree to which mental representations are simplified adapts to the complexity and cognitive demands of a decision environment (Rosch and Lloyd, 1978). In turn, new representations induced by this process result in a more accurate understanding of the environment's causal structure, because new conceptual distinctions are consistent with prior knowledge (Heit, 1994; Lamberts and Shanks, 1997), and are both constrained and validated by observations made during prior decision episodes (Murphy and Medin 1985; Lamberts and Shanks, 1997). Overall, we pose that there is a positive effect of decision experience on decision performance, mediated by a decision maker's cognitive complexity[1].

We test our conjectures in the context of patent evaluations and termination decisions at a large, high-tech Fortune-500 firm. The firm's intellectual property experts, or patent engineers, periodically evaluated all patents in the firm's portfolio and decided whether to renew or terminate them, depending on the patents' forecasted economic prospects. Interviews with the firm's patent engineers confirmed that they made these decisions individually and without receiving any feedback on how the patents eventually performed. As patent engineers accumulate decision experience, any improvements in their decision performance – the degree to which they accurately forecast a patent's future prospects – are unlikely to be due to reinforcement learning, making this setting suitable for testing our theory.

We find support for our predictions. We analysed a corpus of 40,000 written evaluation statements produced by the firm's patent engineers pertaining to approximately 9,000 patent families between 1990 and 2016. We performed causal mapping of the evaluation statements to observe individuals' mental representations (Huff and Jenkins, 2002; Axelrod, 2015) and measured longitudinal changes in representations as patent engineers accumulated decision experience (number of decisions made). We found that cognitive complexity, i.e., the number of conceptual categories and casual relations used by engineers (Gary and Wood, 2011), increased with decision experience. Importantly, this effect was stronger the more causal ambiguity patent engineers experienced. Further, higher cognitive complexity was positively associated with decision performance: patent engineers who used more complex representations made better decisions, as assessed against actual patent commercialization outcomes.

[1] We argue below that increases in cognitive complexity and their positive impact on performance are more likely to occur when decision makers are inexperienced and unaware of which aspects of a decision problem are most relevant, or "uninformed" (Csaszar and Ostler, 2020).

This study contributes to our understanding of experiential learning by addressing the central question of how accumulating experience generates valuable learning (Morris and Moore 2000; Haunschild and Sullivan, 2002). We specify a cognitive mechanism that generalizes the possibility of learning from experience to include contexts where reinforcement learning cannot provide a complete explanation (Glynn et al. 1994; Gavetti, 2005). These contexts are important because decisions with noisy or unobserved feedback are arguably the norm, rather than an exception, in organizational environments (Levitt and March, 1988; Brehmer, 1980; March, Sproull, and Tamuz, 1991; Nelson, 2008; March, 2010).

One implication of our theory is that experiential learning is not only driven by the objective of improving performance as generally assumed, but also by the need to alleviate the cognitive costs that causally ambiguous decisions demand. Hence, experiential learning can be characterized as changes in cognition or behaviour that aim to maximize not decision performance per se, but performance relative to the cognitive costs of decisions. This novel view of experiential learning suggests a wider range of explanations of both individual and organizational level change and adaptation. For instance, it can explain changes that may or may not occur in response to environmental shifts as an attempt to improve or maintain efficiency, especially when it is not immediately clear how performance can be improved (Barr, Stimpert, and Huff, 1992, Eggers and Kaplan, 2009).

Further, our theory shows that noisy or unobserved feedback is not necessarily detrimental to learning as generally assumed. While extant theories argue that decision makers learn "superstitiously" when causal ambiguity is high and feedback is difficult to interpret (Zollo, 2009), the representation learning mechanism shows that these conditions may actually foster learning by inducing the development of more accurate representations.

Finally, our study contributes to the large literature on the role of mental representation in decision-making, which presents conflicting views of the effects of representational complexity (Csaszar and Ostler, 2020). Research has examined the performance advantages of low complexity and fast-and-frugal heuristics (Gigerenzer and Goldstein, 1996; Sull and Eisenhardt, 2015), of highly complex and accurate representations (Kiesler and Sproull, 1982; Weick, Sutcliffe, and Obstfeld, 1999), and of representations that match the complexity of the environment (Ashby, 1956). We inform these views by emphasizing not the performance consequences of different degrees of complexity, but the process by which decision makers are induced to adjust their representational complexity as they accumulate decision experience.

INDIVIDUAL DECISION-MAKING AND LEARNING IN ORGANIZATIONS

A central tenet in the literature on organizational learning is that learning is explicitly or implicitly explained by way of a reinforcement mechanism[2] (Levitt and March, 1988; Levinthal and Rerup, 2006). Decision makers adjust their beliefs and behaviours based on performance feedback, defined as the difference between the outcomes that decision makers observe ex-post and the desired outcomes they intended to achieve ex-ante (March and Simon, 1958; Cyert and March, 1963; Mosakowski, 1997; Gavetti and Levinthal, 2000; Greve, 2003). By implication, reinforcement learning results in improved performance only if performance feedback is observed and furnishes intelligible information about the performance consequences of past actions[3] (Reagans, Argote, and Brooks, 2005; Bae, Biddle, and Park, 2022).

[2] Reinforcement learning is not the only proposed mechanism to describe how learning occurs, yet it is conceptually central (Levinthal and Rerup, 2006). Other learning mechanisms include social influence effects (Bandura, 1977), such as imitation or vicarious learning (March, Sproull, and Tamuz, 1991)

[3] Since we define learning as changes in cognition or behaviour that occur as a function of the accumulation of experience, a discussion of changes in decision performance requires additional definitions. Learning can be related to decision performance only if an appropriate definition of performance is given that specifies: i) the observer who observes outcomes and measures performance; and ii) the procedure used by the observer for assessing and comparing performance before and after changes occur. Subject to these specifications, learning is said to be positive or beneficial (negative or detrimental) from the perspective of the observer if performance is higher (lower) in the final state than in the initial state. This distinction is important because the decision maker and other observers, including the researcher, may use different assessment procedures.

Reinforcement learning may result in superstitious learning when feedback is difficult to interpret, i.e. noisy, with adverse consequences for decision performance (Levitt and March, 1988; Zollo, 2009). Feedback may become noisy for a variety of reasons (Brehmer, 1980; March, Sproull, and Tamuz, 1991; Nelson, 2008; March, 2010). Environmental complexity and stochastic effects may provide spurious evidence of the linkages between actions and outcomes (Sterman, 1989; Denrell, 2008); outcome and feedback signals may be difficult to interpret and to impute to prior actions (Zollo, 2009; Levinthal and Rerup 2021); or feedback may be erroneously attributed to recent actions due to time delays between actions and observed outcomes (Denrell, Fang, and Levinthal, 2004; Rahmandad, 2008). In the extreme, when delays are excessively large or due to information asymmetries, feedback may never be observed (Mosakowski, 1997).

Given that performance feedback is often noisy or unobserved in organizational contexts, we ask whether alternative mechanisms may explain experiential learning, for two reasons. First, empirical observations suggest that improvements of performance do in fact occur in these conditions. Research has documented improvements of decision performance in difficult learning environments such as new product introductions (Paich and Sterman, 1993; Gary and Wood, 2011) and responses to external change (Barr, Stimpert, and Huff, 1992). Experimental evidence shows that task performance can systematically increase even when performance feedback is never observed (Harris and Rosenthal, 1985; Kluger and DeNisi, 1996), and that "individuals will improve their performance on unfamiliar tasks even if they are not given goals and feedback" (Greve 2003: 21).

Second, research in the neuro-cognitive sciences has examined complementary mechanisms to reinforcement learning. Cognitive psychologists have studied processes of categorization and category learning, examining how conceptual categories are acquired and evolve (Rosch et al. 1976; Kruschke, 1992; Murphy, 2004). Relatedly, recent work in the neuro-cognitive sciences has examined the neural bases of representation or structure learning (Tenenbaum et al., 2011; Collins and Frank, 2013; Gershman and Niv, 2013), defined as “the process of learning a useful and compact mapping between observations and states in a specific task” where “usefulness can be measured by how efficiently one can solve a task given the current representation” (Radulescu, Shin, and Niv, 2021: 254). Overall, this work suggests that individuals learn by generating cognitively efficient representations of the environment due to mechanisms that do not exclusively depend on performance feedback (Rosch and Lloyd, 1978; Radulescu, Niv, and Ballard, 2019).

Recent work in organizational learning literature has examined learning processes that address the limitations of reinforcement learning. Deliberate (Zollo and Winter, 2002) and mindful (Weick, Sutcliffe, and Obstfled, 1999; Levinthal and Rerup, 2006) views of learning are central among these contributions[4]. Deliberate learning is attained via knowledge articulation and codification (Cangelosi and Dill, 1965; Zollo and Winter, 2002), allowing decision makers to purposefully invest attention and resources in improving their causal understanding of the environment and thus decision performance (Zollo, 2009). Relatedly, scholars have built on the notion of mindfulness developed in psychology literature (Langer, 1989) to analyse cognitive processes that complement the automaticity associated with reinforcement (Weick and Roberts, 1993; Fiol and O'Connor, 2003). Mindful cognitive states describe “the continual creation and refinement of categories [..] and a willingness to view contexts from multiple perspectives” (Levinthal and Rerup, 2006: 502), which allow individuals to learn from experiences that are difficult to interpret (March, Sproull, and Tamuz, 1991; Weick, Sutcliffe, and Obstfeld, 1999).

[4] They partially integrate or are closely related to equally important notions, such as analogical reasoning (Gavetti, Levinthal, and Rivkin, 2005), cognitive search (Gavetti and Levinthal, 2000; Csaszar and Levinthal 2016), counterfactual thinking (Morris and Moore, 2000), and dialogic practices (Tsoukas, 2009).

However, these contributions do not address the question of how experiential learning can be explained when performance feedback is noisy or unobserved. First, the accumulation of experience is not the main explanatory variable in these theories. The notion of deliberate learning does not suggest that individuals become more likely to make learning investments as they accumulate experience. Similarly, research in mindfulness does not discuss the relationship between increasing stocks of experience and the likelihood of entering mindful cognitive states, or of learning mindfully. Second, generally these theories are implicitly or explicitly premised on the assumption that some performance feedback information is available (Zollo and Winter, 2002; Levinthal and Rerup, 2006).

In summary, existing research on reinforcement learning, as well as deliberate and mindful learning, do not offer a complete characterization of individual learning, particularly when feedback is noisy or unobserved. Below, we develop a model of representation learning that theorizes how learning can occur under these circumstances.

HYPOTHESES

The role of mental representations in individual decision-making and learning

We analyse individual decision-making through the lens of mental representation (Brunswik, 1952; Johnson-Laird, 1983; Barr, Stimpert, and Huff, 1992; Walsh, 1995; Gavetti and Levinthal, 2000; Gary and Wood, 2011). Decision makers have limited cognitive capabilities and need simplified mental representations to process the vast amount of information generated by the environment (Simon, 1991). Simplifications are obtained by ignoring important dimensions of reality (Csaszar and Ostler, 2020) or by aggregating them in broader and less distinctive conceptual categories (Rosch et al. 1976; Martignoni, Menon, and Siggelkow, 2016; Choi and Levinthal, 2022). Mental representations are a critical determinant of strategic choices (Porac et al., 1995; Gavetti, 2005; Eggers and Kaplan, 2013) and can vary substantially across decision makers (Chi, Feltovich, and Glaser, 1981; Ericsson and Smith, 1991; Tanaka and Taylor 1991).

A widely adopted characterization of mental representations consists of a set of concepts and causal relations that represent the causal structure of the environment in an approximate, simplified fashion (Barr, Stimpert, and Huff, 1992; Lamberts and Shanks, 1997; Gary and Wood 2011). Concepts are cognitive categories (Rosch et al. 1976; Murphy, 2004) that determine which information from the decision environment receives attention and what is dismissed (Nisbett and Ross, 1980; Kiesler and Sproull, 1982). Relations between concepts allow for more complex information processing than the simple assignment of raw information to categories of meaning (Thagard, 2005; Gopnik and Schulz 2007). Together, concepts and causal relations co-determine how decision makers process stimuli and respond with solutions, opinions, decisions, and actions (Nisbett and Ross, 1980; Dutton and Jackson, 1987).

From the vantage point of mental representation, and in keeping with the definition of learning as changes in cognition or behaviour, we focus on changes in the number and structure of concepts and relations (Bartunek, 1984; Walker, 1985; Lurigio and Carroll, 1985; Walsh, 1995; Denrell, Fang, and Levinthal, 2004) and in how they reflect the causal structure of the environment (Weick, 1979; Barr, Stimpert, and Huff, 1992).

A model of representation learning in individual decision-making

We propose a learning mechanism that we call *representation learning*. Inspired by cognitive categorization theory (Rosch et al., 1976; Lamberts and Shanks, 1997; Rehder, 2003; Murphy, 2004), we hypothesize that the effect of accumulating decision experience on decision performance is mediated by changes in the structure of mental representations. Specifically, we argue that individuals develop more distinctive and fine-grained mental representations of the environment as they repeatedly make decisions under causal-ambiguity and in condition of noisy or unobserved feedback (**H1a, H1b**). In turn, more fine-grained mental representations help individuals make better decisions even in these adverse learning contexts, as they reflect increasingly accurate representations of the environment (**H2**).

--- INSERT FIGURE 1 ABOUT HERE ---

We define cognitive complexity as an increasing function of the number of concepts and relations of mental representations (Simon, 1962; Dane, 2010). For clarity, greater cognitive complexity is not necessarily associated with enhanced representation accuracy, as concepts and interdependences may be unrepresentative of the environment's causal structure (Gary and Wood, 2011) or irrelevant for a given class of decision problems (Csaszar and Ostler, 2020).

We argue that the relationship between decision experience and cognitive complexity is a consequence of the fact that individuals tend to adapt the complexity of their representations to the cognitive demands of the decision environment. This occurs because the structure of mental representations reflects the principle of cognitive economy (Rosch et al. 1976; Rosch and Lloyd, 1978; Murphy and Brownell, 1985; Lamberts and Shanks, 1997; Murphy, 2004). This principle states that human beings' goal is to use representations that "provide maximum information with the least cognitive effort […] conserving finite resources as much as possible" (Rosch and Lloyd, 1978: 28). It follows that the conceptual categories used by individuals reflect an equilibrium resulting from a trade-off between distinctiveness and aggregation (Rosch et al., 1976).

On the one hand, the informative content of categories can be maximized by using finer-grained, highly distinctive concepts that discriminate observations with respect to a large number of dimensions. In the extreme, every observation would be assigned its own concept and stored in memory as a separate entity. However, this degree of distinctiveness would soon exhaust available memory and become excessively costly in terms of retrieval and processing (Rosch et al., 1976). On the other hand, cognitive resources can be preserved by way of coarsergrained, simplified concepts that aggregate observations by neglecting contingencies and differences across dimensions. (Gershman and Niv, 2013). However, excessively simplified concepts “are harder to use, because it is difficult to tell them apart” (Murphy, 2004: 219). Excessive aggregation may increase rather than decrease cognitive efforts because it makes it difficult to disambiguate the relevant dimensions of a decision problem and identify effective courses of action.

Thus, individuals tend to adopt representations that balance these two conflicting tendencies – that is, they are induced to optimize the number of dimensions and contingencies that conceptual categories can encode relative to the cognitive resources they demand. This insight allows us to reflect on how cognitive complexity evolves as individuals accumulate decision experience, i.e., repeatedly make decisions of a similar nature over time.

We argue that the equilibrium between distinctiveness and aggregation is altered towards greater distinctiveness, and thus greater complexity, when individuals repeatedly make decisions under causal ambiguity. The subjective causal ambiguity perceived by decision makers is the dimension of decision uncertainty that is most relevant in this context, defined as the extent to which action-outcome linkages are understood and perceived as clear by decision makers (Konlechner and Ambrosini, 2019). Perceptions of causal ambiguity require individuals to deploy cognitive work, time and time again, aimed at identifying effective courses of action. Given that the reduction of these costs is a “fundamental need” of individuals (Hogg and Mullin, 1999: 253), individuals are induced to increase distinctiveness when causal ambiguity is persistent in order to restore cognitive efficiency. This is expected to occur until further increases in the cognitive load associated with processing more complex representations exceed the costs associated with causal ambiguity (Finton, 2005).

We note that increases in cognitive complexity are expected to occur under the following two general conditions. First, individuals tend to increase complexity when they are not yet aware of most relevant dimensions and contingencies of the environment, i.e. they are “uninformed” (Csaszar and Ostler, 2020: 9). Second, increases tend to occur for repeated decisions that are not prescribed by rules or standard operating procedures and over which decision makers have considerable discretion - i.e. decisions for which causal ambiguity is substantial and persistent (Mosakowski, 1997; Michel, 2007). While conditions may exist under which individuals are induced to simplify representations, the above conditions ensure that complexity tends to increase[5], especially when performance feedback is systematically difficult to interpret or observe. In these conditions, decision makers cannot rely on trial-and-error iterations to understand which dimensions are most relevant or identify effective courses of action. Further, they have no reasons to believe that simplifying representations by focusing on a subset of dimensions or actions would increase their chances to achieve desired outcomes, because they still do not know which dimensions are most relevant or actions are most effective among the ones they are aware of. Rather, perceptions of causal ambiguity might even increase by simplifying representations, because decision makers would be aware that they are omitting potentially relevant dimensions and interdependencies that they have previously taken into account.

Whereas feedback is central to reinforcement learning, the above line of reasoning explains changes in mental representation solely as a function of the repeated exposure to causal ambiguity, regardless of whether any performance feedback is observed. Therefore, we pose:

H1a: *As individuals accumulate experience in making decisions in a domain, their mental representations relating to this domain become more complex*.

If the development of more complex mental representations is a function of repeated exposure to decisions that are causally ambiguous, we should expect this effect to be more pronounced,

[5] These conditions are especially important because we are interested in studying how individuals learn from their own experience when they do not possess pre-existing knowledge of the decision environment and when their decisions are not guided by knowledge embedded in the organization (March and Simon, 1958)

the greater the extent of causal ambiguity the decision maker perceives. The higher and more persistent the causal ambiguity perceived during repeated decisions, the greater the efforts and the need to adjust representations to optimize cognitive efficiency. Accordingly, we predict:

H1b: *The positive effect of accumulating decision experience on cognitive complexity is strengthened by the cumulative degree of causal ambiguity experienced.*

The relationship between cognitive complexity and decision performance is a consequence of the fact that the above process induces changes in cognition that are accurate representations of actual dimensions and interdependencies of the environment.

New finer-grained distinctions are accurate for two reasons. First, new distinctions are derived and inherit accuracy from previously held concepts and relations. We can build again on the principle of cognitive economy and on further results from cognitive psychology literature. Any conceptual change that occurs to satisfy the requirements of cognitive economy is likely consistent with prior concepts and relations in order to avoid additional ambiguities and inconsistencies (Fiske and Taylor, 1991). Changes that conflict with what is already known would increase rather than alleviate cognitive efforts. This phenomenon is documented in the cognitive psychology literature and described in terms of integration (Heit, 1994) and selective weighting (Murphy and Medin, 1985) effects of prior knowledge on the generation of new concepts and relations (Lamberts and Shanks, 1997).

Second, new concepts and relations are derived from previous observations of the actual functioning of the environment's causal structure, rather than as acts of pure imagination. In their seminal work on the effects of prior knowledge on concept learning, Murphy and Medin argue that individuals form "background beliefs" about the decision environment that constrain the generation of new concepts and relations (Murphy and Medin 1985: 303). They argue that representations are cognitive devices that reflect users' needs and decision objectives, but regardless of how simplified representations are, individuals are still exposed to the multiplicity of dimensions and interdependencies that characterize an environment. In fact, background beliefs are raw data that individuals store in memory as they accumulate experience and observe how cause-effect chains of events unfold in the environment. These data contain information about causal relations and thus about what is possible and what is unlikely to occur in a domain. As such, background beliefs restrict the "space of hypotheses" that individuals would consider as plausible for conceiving new concepts and relations[6] (Lamberts and Shanks 1997: 12). It follows that new distinctions are constrained by previous observations that are stored in memory and gain accuracy from them.

In turn, research on managerial cognition shows that accurate representations improve decision performance (Bourgeois, 1985; Barr, Stimpert, and Huff, 1992; Gary and Wood, 2011). Accurate representations provide a more complete and deeper understanding of causes and effects, allowing decision makers to make more precise estimates of the distributions of outcomes and thus to choose more effective actions (Einhorn and Hogarth, 1986). For instance, Lurigio and Carroll's (1985) study of probation officers demonstrated that respondents with higher representational accuracy made higher quality decisions and processed information more easily and confidently; while McNamara, Luce and Tompson (2002) showed that the complexity of top managers' representations of the competitive environment was associated with higher firm performance. Gary and Wood's (2011) managerial simulation study provided experimental evidence of the performance benefits of representational accuracy. Participants made strategic decisions in a simulated environment to maximize profits from product sales. The authors used answers to standardized tests to observe participants' mental representations during the experiment. The results show that more accurate mental representations improved decision performance (Gary and Wood, 2011).

[6] For instance, in the patenting context, it would be unreasonable to expect that patents' value depends on the day of the week in which a patent application was filed or on that day's weather. The fact that this independence would be obvious to intellectual property experts, and even to non-experts, is precisely the consequence of possessing background beliefs, which constrain the range of plausible hypotheses.

As previously discussed, it is important to note that we expect increases in complexity to be more likely beneficial to performance when individuals are relatively inexperienced and uninformed of the most relevant dimensions of the environment. Consistent with the patterns observed by Bingham and Eisenhardt (2011) and with the results obtained by Csaszar and Ostler (2020), our argument suggests that decision performance increases with complexity and accuracy in the more general case in which we do not assume pre-existing levels of experience or of knowledge of the casual structure. It is in these cases that individuals cannot discriminate among important contingencies and courses of action and that representations are likely too simple or "underspecified" (Martignoni, Menon, and Siggelkow, 2016: 2545). Therefore, we pose:

H2: *Increases in cognitive complexity relating to a decision domain are associated with higher decision performance.*

In sum, our representation learning mechanism states that the accumulation of decision experience drives greater cognitive complexity, which is in turn associated with higher decision performance. We pose:

H3a: *Cognitive complexity relating to a decision domain mediates the positive relationship between the accumulation of decision experience and decision performance.*

H3b: *The positive relationship between decision experience and decision performance mediated by cognitive complexity is strengthened by the cumulative degree of causal ambiguity experienced.*

EMPIRICAL CONTEXT: MANAGING THE PATENT PORTFOLIO AT ALPHA

To test our hypotheses, we require a context where individuals in an organization: (a) repeatedly make decisions in a specific domain; and (b) receive noisy or no feedback.

Accordingly, we study decisions relating to the recurring evaluation of patents within the patent portfolio at Alpha, a multinational, Fortune-500 ITC firm (pseudonym).

Many large firms regularly review all active portfolio patents in the to identify opportunities for value creation while saving on maintenance costs. The regular re-evaluation of patent portfolios is necessary because the value a firm attributes to a patent changes over its lifetime, as new competing technologies are developed or new market opportunities arise (Guler, 2007; Khanna, Guler, and Nerkar, 2018). Patent rights can be renewed with the respective patent offices for up to 20 years, subject to the payment of recurring maintenance fees. Maintenance costs can be reduced by terminating patents, frequently by reducing the size of a given patent family rather than by terminating the whole family at once. Serrano (2010) estimates that nearly 50% of all patents are terminated before their legal term by their owners.

In order to better understand the patent evaluation process at Alpha, we conducted interviews with employees and with Alpha's director of IP, and we obtained access to and analysed evaluation guidelines and other internal documents. We conducted one-hour interviews with 16 patent evaluators based in 7 locations worldwide and collected 10 responses from evaluators to a 15-question survey. This qualitative evidence confirmed the suitability of this empirical setting for testing our theory.

At Alpha, regular patent evaluations were performed by patent engineers. They could solicit advice from Alpha's technology experts, including patent inventors, as deemed necessary, but remained solely responsible for making termination decisions. Patents were typically re-evaluated over time by the same patent engineer, although they were occasionally reallocated to a different patent engineer, due to personnel mobility, for instance.

For a patent engineer, patent evaluation consisted of writing an evaluation statement, assigning a rating, and terminating one or more family members as deemed necessary. The statements were meant to describe a patent's limitations and highlight opportunities for value creation that the firm's business units could potentially exploit. Additionally, the statements needed to provide useful information for future re-evaluations. Accordingly, patent engineers were required to produce exhaustive evaluation statements that articulated their reasonings and included any relevant information for future reference.

Ratings were numerical ranging from 0 (low) to 5 (high) according to broad evaluation guidelines. Patent engineers could also add identifiers to the numerical ratings to identify patents that could potentially become part of a technology standard or be implemented in Alpha's products. The evaluation history of all patents was stored in a software system to which patent engineers had unrestricted access. The system was also accessible to Alpha's commercialization units, such as standardization, infringement and litigation, or the product implementation units. These units used ratings to identify higher potential cases and the information provided by the evaluation statements to guide their commercialization efforts.

The suitability of this setting for studying learning in conditions of noisy or unobserved feedback is underscored by the substantial causal ambiguity to which patent engineers were exposed and by the extended time lags between evaluations and commercialization outcomes. The fact that evaluations were made in conditions of substantial causal ambiguity emerged from interviews we conducted with patent engineers and from an analysis of internal evaluation documents and guidelines. Alpha did not provide specific criteria or rules for making evaluations, and patent engineers had significant discretion over ratings and terminations. Internal documents and evaluation guidelines mention broad dimensions or evaluation factors against which patents' future economic prospects could be evaluated, such as 'legal protection' or 'business value'. However, Alpha did not provide training or instructions on how the specific characteristics of patents could be assessed and mapped into those broad factors. Patent engineers relied instead on their own knowledge and subjective judgement. They admitted that it was often difficult to identify factors that could clearly indicate whether patents had potential for commercialization. Engineers also confirmed that these patents were given intermediate ratings (2 and 3) and that were the "*most difficult to evaluate*". This type of causal ambiguity reflects the longstanding challenges of assessing patent value, as discussed in the patent literature (Wang and Hsieh, 2015; Higham, de Rassenfosse, and Jaffe, 2021).

Extended time lags between evaluations and commercialization outcomes made it particularly challenging to learn from the outcomes of past decisions in this context. Commercialization outcomes frequently materialize several years after evaluations are conducted. Larger time lags increase the likelihood that exogenous changes, such as technological, legal, or competitive shifts, interfere with the realization of commercialization efforts, adding significant noise to feedback signals.

Most importantly, extended time lags also meant that patent engineers typically could not observe the outcomes of their decisions. This can be attributed to three reasons. First, ongoing commercialization efforts and eventual (un)successful outcomes were not internally communicated to patent engineers at Alpha. Patents commercialized by the firm's business units were simply removed from the list of patents allocated to patent engineers, and units were not instructed to notify patent engineers of initiated or finalized attempts to commercialize patents they had previously evaluated. Indeed, all interviewees confirmed that, in the vast majority of cases, they "*do not know what happens to the patents*" they evaluate.

Second, patents were often reallocated to different patent engineers, a practice that further decreased the limited opportunities that they might have had to observe the outcomes of their evaluations. Patent engineers may have changed role within Alpha or left the firm during the period between their evaluations and the realization of commercialization outcomes.

Finally, even if Alpha's business units provided information about outcomes, patent engineers would have likely struggled to recollect why and how they had used certain evaluation factors several years in the past, as suggested by the following interview quote: "*You don't know if you* *do good or bad. And when you know, it's too much in the past*". Large delays significantly disrupt the learning value of feedback because individuals may not be able to link feedback information to the problem representations that guided their decisions (Rahmandad, 2008; Denrell, Fang, and Levinthal, 2004).

Therefore, in our setting, systematic improvements of decision performance are unlikely to be explained by reinforcement learning, that is, in terms of an iterative adjustments to performance feedback. Patent evaluation at Alpha thus offers a suitable context in which to explore experiential learning mechanisms that do not depend on feedback.

DATA AND METHODS

Data

We constructed a dataset with information drawn from Alpha's patent portfolio management system, matched with archival data on patents obtained from external sources. We collected a corpus of 40,000 written evaluation statements pertaining to approximately 9,000 patent families that Alpha filed between 1988 and 2006. Most of the evaluations were made within the first ten years of patents' life (92%), and each family was evaluated nearly three times on average. The statements were produced by nearly 146 Alpha patent engineers, who each, on average, evaluated 161 patents and made 450 evaluations between 1990 and 2016. The average evaluation statements contained 101 words.

We used Alpha sources and external sources to collect information about successful patent commercialization outcomes, which include licensing agreements, sales, patent standardizations, litigations, and implementations in products. Alpha's commercialization units recorded dates and details of commercialization events. We complemented and validated this information with data collected from two external databases. We collected data on transfers of ownership and litigation legal actions from Google Patents, and additional data on litigation legal actions and infringements from Clarivate's Darts-IP database. More than 3,700 (23%) patents were successfully commercialized during the observation period. Commercialization occurred with approximately equal frequency between two and fifteen years after filing, while we observed a sharp decline of the rate of commercialization events for older patents (greater than sixteen years).

Variables and Measurements

Decision performance

The variable describes the extent to which patent engineers correctly estimated the future value of patents and accounts for both false positives (Type I errors) and false negatives (Type II errors). We used changes in engineers' patent family ratings as indicative of positive or negative evaluations. The guidelines recommended, but did not prescribe, reducing or "*trimming"* the number of active members for a patent family with lower ratings (ratings 0 and 1), and renewing or expanding active members for families with higher ratings (ratings 4 and 5). Additionally, as previously mentioned, ratings guided commercialization units' search for valuable patents in the portfolio. Hence, ratings were meaningful indicators of the value that patent engineers attributed to patents, and ratings' increases or decreases reflected patent engineers' positive or negative forecasts of patents' future prospects[7].

Considering changes in ratings, we defined *decision performance* in the following way. Decision performance is positive when patent engineers increased or maintained the ratings of patents they believed had positive future prospects and that were subsequently commercialized; and patent engineers decreased the ratings of patents they believed had negative future prospects and that subsequently expired without being commercialized. Conversely, decision

[7] The average rating assigned to a patent was 2.54 (SD 1.01). Intermediate ratings 2 and 3 were assigned respectively, 42% and 38% of the times and ratings 1 and 4 were assigned respectively, 6% and 8% of the times. Nearly 65% of all evaluations confirmed the most recently assigned rating, whereas ratings were upgraded or downgraded by one unit respectively, 16% and 11% of the times and by two units nearly 4% and 3% of the times. Most upgrades by one unit occurred for families rated 2 (58%) and 3 (24%), while downgrades by one unit occurred for families rated 3 (55%) and 2 (27%).

performance is negative when patent engineers increased or maintained the ratings of patents they believed had positive future prospects and that subsequently expired without being commercialized (false positive); and when patent engineers decreased the ratings of patents they believed had negative future prospects and that were subsequently commercialized (false negative). See Appendix A for an example of how we computed decision performance.

Our measure of decision performance has at least three desirable characteristics. It is an objective measure of the accuracy of patent engineers' estimates (Zollo, 2009). Second, the measure accounts symmetrically for both the objectives of the patent evaluation process at Alpha. That is, it captures the accuracy of evaluations in terms of identifying both high value patents for commercialization and low value patents for saving on maintenance and management costs. Finally, the measure avoids issues of manipulability (Zollo, 2009), because commercialization decisions did not depend on patent engineers but on the business units and other parties.

As a caveat, we cannot measure the accuracy of decisions to terminate all the active members of a patent family, because it is not possible to know whether commercialization could have occurred had the patent engineers not decided to abandon these families. However, this issue has limited consequences in our setting because abandonment decisions were made for less than 6% of the families in the portfolio.

We operationalize *decision performance* as an ordered categorical variable defined as:

$$performance = (assigned\ rating - previous\ rating)\ \times commercialization$$

The variable $commercialization$ is equal to 1 if patents were eventually commercialized and equal to -1 if patents expired without being commercialized, while *assigned rating* and *previous rating* were, respectively, the rating assigned by the patent engineer as a result of the focal evaluation and the rating assigned to the patent as a result of the most recent evaluation[8].

Cognitive complexity

This variable describes the number of concepts and causal relations that characterize patent engineers' mental representations. In our context, concepts and relations correspond to patent engineers' understanding of evaluation factors, such as 'novelty', 'potential for standardization', and 'scope of legal claims', and of how these factors were interrelated. Cognitive complexity increases with both the number of concepts that represent evaluation factors and the number of interdependencies among them (Simon, 1991; Dane, 2010).

We observed the evolution of patent engineers' mental representations over time by applying a causal mapping method to the evaluation statements they produced. Causal mapping is a form of content analysis, whose purpose is to identify and categorize concepts and causal relationships within a document (Barr, Stimpert, and Huff 1992; Huff and Jenkins, 2002; Gary and Wood, 2011; Axelrod, 2015).

We proceeded as follows. We performed an iterative coding procedure, which yielded 28 evaluation factors that patent engineers collectively used over the entire observation period (Appendix B). We initially applied topic modelling (Hannigan et al. 2019) and word embedding methods to the evaluation statements and obtained a first list of factors. We validated and refined this list using rating descriptions contained in the evaluation guidelines and analysed patent engineers' email exchanges with inventors and domain experts. Patent engineers often solicited evaluation advice from inventors and experts by sending forms containing standard sets of questions on certain topics. We used these topics to further refine

[8] Each patent entered Alpha's portfolio with a preassigned rating.

our list of factors. Eventually, this procedure yielded 28 evaluation factors and, for each factor, a dictionary of keywords and key phrases.

We then mapped the content of each evaluation statement against the factors and causal connections between factors within the text. Using dictionaries of keywords and key phrases, we matched each phrase contained in the statements to an evaluation factor. We then identified causal connecting words between phrases associated with different factors to estimate the extent to which patent engineers made causal connections between factors[9]. To illustrate, we show the computation of complexity for two sample evaluation statements in Table 1.

--- INSERT TABLE 1 ABOUT HERE ---

We computed a patent engineer's mental representation complexity by counting the distinct evaluation factors and causal connections that occurred in the evaluation statements they produced in a given time window[10]. Our measure of complexity can be expressed as follows:

$$complexity_{t,t-t^*} = (evaluation\ factors)_{t,t-t^*} + (causal\ relations)_{t,t-t^*}$$

where the time periods $t, t - t^*$ define the observation window prior to the time of evaluation; and *evaluation factors*, *causal relations* are, respectively, the average number of distinct evaluation factors and causal connecting words occurring in the evaluation statements produced by a patent engineer during the 10 most recent evaluations prior to the focal evaluation.

In robustness tests, we define alternative windows to include the 5, 20, or 30 most recent prior evaluations; or 10, 30, or 60 days prior to the focal evaluation. Additionally, we use measures of cognitive complexity that only accounts for the number of causal connections that occur in the evaluation statements (Gary and Wood, 2011) and measures that are weighted by the word

[9] We used the dictionary of causal keywords defined in the LIWC software, such as *thus* and *because*.
[10] Multiple evaluation statements written by a same patent engineer over a short period of time must all reflect the same mental representations held by the patent engineer during that time. Mental representations change gradually and are relatively persistent (Walsh, 1995), and hence the cognition of a patent engineer will not substantially change between two evaluations made over a sufficiently short period of time.

length of the evaluation statements. These sets of measures provide results that are consistent with our chosen measure.

Causal ambiguity

This variable describes the causal ambiguity to which patent engineers were exposed during evaluations prior to the focal evaluation. In line with strategic decision-making literature (Mosakowski, 1997), we measure *causal ambiguity* as a linearly decreasing function of patents' age. *Age* measures the difference between the family's earliest filing year and the year at the time of evaluation. We expect age to be inversely related to the residual causal ambiguity perceived by patent engineers because information on future value creation opportunities is disclosed over time and, concurrently, the likelihood of positive outcomes decreases over time, (Mosakowski, 1997; Konlechner and Ambrosini, 2019). Since we are interested in capturing the degree to which perceived causal ambiguity was persistent over time, we operationalize *causal ambiguity* as a linearly decreasing function of the average patents' age of the 10 most recent prior evaluations, which we scale so that it ranges from 0 (low) to 1 (high).

We define alternative operationalizations in robustness tests. We expect perceived causal ambiguity to be lower for patents that were rated relatively very low (0 and 1) and very high (4 and 5). In our interviews, respondents confirmed that patents rated 2 or 3 were the most uncertain and thus difficult to evaluate. Accordingly, we alternatively compute *causal ambiguity* as the proportion of patents with previous rating 2 and 3 among the 10 most recent prior evaluations. The variable ranges from 0 to 1. We also define alternative windows for both this and our chosen measure to include the 5, 20, or 30 most recent prior evaluations; or 10, 30, or 60 days prior to the focal evaluation. These alternative operationalizations produce results that are consistent with those presented here.

Decision experience

This variable simply counts the number of decisions made by a patent engineer prior to a focal evaluation. Patents were randomly allocated to patent engineers. For instance, interviewees confirmed that patents were not allocated on the basis of patent engineers' seniority or evaluation competence. Rather, the number and type of allocated patents depended on circumstances, such as incoming renewal deadlines, specific needs related to ongoing R&D projects, or periodic budget constraints.

Control variables

We included control variables pertaining to both decision and patent family levels. First, we included variables to account for different features across decisions. To account for external input that patent engineers received to reach a decision, we included two dummy variables, *inventor opinion* and *expert opinion,* which indicate whether the focal patent engineer obtained opinions from an inventor (of the patent to be evaluated) and from other Alpha experts, respectively. We also constructed a dummy variable, *transferred case*, which accounts for whether previous evaluations of the focal patent were performed by another patent engineer. While *causal ambiguity* accounts for the average causal ambiguity to which patent engineers were exposed during the most recent prior evaluations, we included the *causally ambiguous patent* variable to account for the perceived casual ambiguity related to the focal evaluation. We operationalize this variable as a linearly decreasing function of the focal patent's age, which we scale so that it ranges from 0 (low) to 1 (high). Similarly, we included the dummy variable, *flagged patent*, which equals 1 if the focal family was identified as potentially relevant to technology standards or product implementation in Alpha prior to the focal evaluation. To control for attention and cognitive capacity, we computed the number of evaluations made by the focal patent engineer in the previous seven days (*workload*). Finally, we include the *word length* of the focal evaluation statement.

Second, in line with previous work (Khanna, Guler, and Nerkar, 2018; Higham, de Rassenfosse, and Jaffe, 2020), we included several patent family level variables that could influence patent engineers' perception of future patent value. *Family size* and *number of claims* can be associated with the scope of patent protection and, thus, with patent value. *Family size* is the number of jurisdictions where patent applications were submitted, while *number of claims* is the maximum number of independent claims across the patents of the family. We counted the maximum number of *forward* and *backward citations* across the family to account for potential signals of quality. Further, the number of granted and abandoned family members at the time of evaluation could have respectively been perceived as positive and negative quality signals by patent engineers. Accordingly, *granted ratio* and *abandoned ratio* are the proportions of respectively granted and abandoned family members to the family size at any time up to the evaluation. We include *filing year* dummy variables to control for variations in patent quality due to time trends and 16 *technology* dummy variables based on the internal technological classification to account for differences across technologies.

Empirical strategy

To test Hypotheses 1a and 1b, we used panel data linear models adjusted for patent engineer fixed effects, filing year and technology dummies, and robust standard errors. The models represent the path *a* between decision experience and complexity and the interaction of causal ambiguity between the two variables.

To test Hypotheses 2 and 3a/b, we conducted a moderated mediation analysis using generalized structural equation modelling (GSEM) with bootstrapping (Edwards and Lambert, 2007; Preacher, Rucker, and Hayes, 2007; Hayes, 2013). We used panel data-ordered logit and linear models, adjusted for patent engineer fixed effects, filing year and technology dummies, and robust standard errors. Moderated mediation exists when a moderator variable affects the path between independent and mediator variables. In our model, the path *a* between the independent variable (decision experience) and mediator variable (cognitive complexity) depends on the value of the moderator (casual ambiguity, Hypothesis 1b). The path *b* between the moderator and the dependent variable (performance) does not depend on the value of the moderator (Hypothesis 2). The mediated effect – or *Indirect Effect,* path *ab* – is alternative to the *Direct Effect* – path *c* – of the independent variable on the dependent variable. The *Conditional Indirect Effect* or CIE is the value of the *Indirect Effect* conditional on the value of the moderator variable. The sum of the CIE and of the *Direct Effect* is equal to the *Total Effect* *c'=c + ab*. Mediation (Hypothesis 3a) is tested by verifying the statistical significance of the CIE and of the *Direct Effect*. Full mediation occurs when the *Direct Effect* is not significant. Otherwise, the degree of partial mediation is estimated by comparing the CIE to the *Total Effect*[11].

We used standardized variables for decision experience and causal ambiguity in all models. Further, we ran all models on the subsample of patent engineers who performed at least 15 patent evaluations, consisting of 146 out of nearly 180 patent engineers in our full sample.

RESULTS

Table 2 and 3 report summary statistics and pair-wise correlations for the variables used in our models. Most of the correlation coefficients are low. We derived variance inflation factors (VIF) for all models (Greene, 2003). All the computed values were less than 3 and the mean VIF was less than 1.6, indicating that multicollinearity is not a concern in the regressions.

--- INSERT TABLE 2 & 3 ABOUT HERE ---

Hypotheses 1a and 1b are supported. Table 4 shows the results for panel data linear models that we used to test Hypothesis 1a and 1b. We predicted a positive relationship between decision experience and cognitive complexity (H1a) and a positive moderating effect of causal

[11] The conditional indirect effect is calculated by estimating the following two equations, where X is the independent variable, M is the mediator variable, W is the moderator variable, and Y is the dependent variable: $M = a_0 + a_1X + a_2W + a_3XW$ and $Y = b_0 + b_1M + b_2X + b_3W + b_4XW$.

ambiguity (H1b). Model 2 indicates a positive and significant relationship between experience and complexity ($B = 0.16, p < 0.01$). Model 3 indicates a positive and significant interaction between experience and causal ambiguity ($B = 0.07, p < 0.01$).

These results show that one standard deviation increase in decision experience – or about 122 evaluations - is associated with 13% and 37% of one standard deviation increase in complexity when causal ambiguity is one standard deviation below and above its mean value, respectively. The average complexity for relatively inexperienced patent engineers who made 25 or less evaluations (lower quartile of *decision experience*) is nearly 2, while average complexity for engineers who made 300 or more evaluations (10% of all patent engineers) increases by more than 50%. Further, we verified in additional analyses that increases in complexity result from changes in representations that are incremental and consistent with previously held representations[12].

The estimates for the other controls are consistent across specifications and confirm our expectations. The workload allocated to the patent engineers at the time of evaluation is negatively associated with cognitive complexity, indicating that accounting for more factors and interdependencies demands more attention and cognitive resources. A negative coefficient for *flagged patent* was also expected because these patents were likely evaluated with evaluation factors that were predominantly related to standardization and implementation. Cognitive complexity also tends to be lower when patent engineers requested evaluation advice from the inventors (*inventor opinion, p<0.05*) and from Alpha experts (*expert opinion, p>0.1*). Similarly, complexity is lower when patent engineers have access to the evaluation statements of previous patent engineers made for transferred patents (*transferred case*). These results

[12] Evaluation factors that patent engineers have used more frequently in the past are significantly more likely to be used in the future as compared to less frequently used ones, and tend not to be replaced by new factors. Thus, increases in complexity result from new evaluation factors that patent engineers start using as they accumulate experience and that stabilize over time; and from new causal relations between factors.

suggest that increases in the complexity of representations is unlikely to be explained by the external influence of the opinions provided by inventors, experts, or other patent engineers.

--- INSERT TABLE 4 ABOUT HERE ---

Table 5 and 6 report the results for moderated mediation GSEM analyses with bootstrapping that we used to test Hypothesis 2 and 3a/b. Additionally, the results provide further evidence in support of Hypothesis 1a and 1b. We predicted a positive relationship between cognitive complexity and decision performance (H2) and the mediating effect of complexity (H3a) conditional on the effect of causal ambiguity (H3b). In Table 6, the results for the bootstrapping confidence intervals confirm, as above, that the effect of experience on complexity (path a) and the moderating effect of causal ambiguity are positive and significant.

Bootstrapping confidence intervals also confirm that the effect of cognitive complexity on decision performance (path b) is positive and significant, indicating support for Hypothesis 2 ($B = 0.07$, $p < .01$). Additionally, Table 5 shows ordered logit models adjusted for patent engineer fixed effects, filing year and technology dummies. The models confirm Hypothesis 2 ($B = 0.07, p < 0.01$).

Hypothesis 3a predicted that the relationship between decision experience and performance is mediated by cognitive complexity. The results for the bootstrapping confidence intervals indicate that the direct effect of experience on performance (path c) is not significant ($B = 0.06$, $p > 0.1$), while the CIE is positive and significant for low, mean, and high values of causal ambiguity (respectively one standard deviation below mean prior uncertainty ($B = 0.01$, $p < 0.01$); mean prior uncertainty ($B = 0.01$, $p < 0.01$); and one standard deviation above mean prior uncertainty, $B = 0.02$, $p < 0.01$). It follows that cognitive complexity fully mediates the relationship between experience and performance, indicating support for Hypothesis 3a.

These results also confirm Hypothesis 3b. We predicted that the positive indirect relationship between decision experience and performance is strengthened by the cumulative degree of causal ambiguity to which patent engineers were exposed. We tested this hypothesis by computing the difference between the CIEs at different values of prior uncertainty, as shown above (Hayes, 2013; Edwards and Lambert, 2007). The results for the bootstrapping confidence intervals of these differences indicate support for Hypotheses 3b. In particular, the difference between the CIEs at mean and low levels of prior uncertainty, and between the CIEs at high and mean levels of prior uncertainty, are positive and statistically significant ($CIE_{mean} - CIE_{mean-sd} = 0.005, \quad p < 0.01,$ and $CIE_{mean+sd} - CIE_{mean} = 0.005, \quad p < 0.01,$ respectively).

--- *INSERT TABLE 6 ABOUT HERE* ---

The estimates for the other controls are consistent across specifications and mostly confirm our expectations. The workload allocated to the patent engineers at the time of evaluation is negatively associated with performance. On the contrary, families that were flagged as potentially relevant for standardization or implementation (*flagged patent*) are significantly and largely associated with higher performance. Indeed, this information is especially valuable in terms of reducing causal ambiguity and facilitating patent engineers' forecasts. Finally, the opinions of the inventors seem to bias rather than inform patent engineers. This effect may be due to the inventors' overly positive assessments of their own creations. In additional analyses we verified that the availability of inventors' opinions is in fact positively and largely associated with likelihood of Type I errors.

DISCUSSION AND CONCLUSIONS

In this paper, we explore how patent engineers in a large corporation learned how to make patent evaluations and termination decisions. We show that the quality of patent engineers' decisions improved as they gained experience in evaluating the commercial prospects of patents, even though they almost never obtained feedback on how their decisions fared. We argue that this happens because these decision makers develop more complex mental representations as they repeatedly engage in decisions of a specific class. Accordingly, we found that individuals who had made more decisions predicted actual patents outcomes more accurately; and that this effect was mediated by the cognitive complexity individuals had developed as a result of accumulating experience. We further found that the positive effect of decision experience was accentuated when decisions were persistently perceived as causally ambiguous. We refer to the type of feedback-less learning as representation learning.

Our study has implications for three bodies of literature. Our first contribution is to prior work on experiential learning. The experiential learning framework is premised on the beneficial effects of accumulating experience in a decision environment (Levitt and March, 1988). The main explanatory variable is the stock of experience gained by a learning actor and the central explanation is the reinforcement mechanism (Argote and Miron-Spektor, 2011).

Representation learning extends the way we conceptualize experiential learning in the following ways. While in reinforcement learning the main function of experience is to provide successive feedback loops between the environment and the decision maker, in representation learning, experience enables a developmental cognitive process that works regardless of whether feedback is observed. This has implications for the boundary conditions under which effective experiential learning can occur, most obviously by expanding the spectrum of situations to include those where feedback is noisy or unobserved. The latter kind of condition is pervasive in organizational contexts (Brehmer, 1980; March, 2010).

Importantly, representation learning also introduces a novel perspective on why learning happens in the first place. One of the central assumptions of reinforcement learning is that decision makers engage in trial-and-error iterations to improve decision performance. We argue that learning is also driven by the need to alleviate perceptions of causal ambiguity. This alternative view is consistent with the definition of learning as changes in cognition or behaviour that occur as a function of experience; and supported by the core tenet of cognitive science literature that individuals need to economize on cognitive resources. It follows that experiential learning can be more broadly characterized as driven by the objective of optimizing decision performance relative to the costs of making decisions. This implies that when it is difficult to verify whether alternative actions or solutions would improve performance, individuals may – or may not – adjust their beliefs or behaviours depending on efficiency considerations. This perspective extends our understanding of managerial responses to environmental change (Barr, Stimpert, and Huff, 1992) and of decision structures aimed at promoting decision efficiency (Simon, 1947) and avoiding uncertainties (Cyert and March, 1963) in organizations.

Further, our theory shows that noisy or unobserved feedback is not necessarily detrimental to learning as generally assumed. Extant theories argue that decision makers learn spurious action-outcome associations when feedback is largely delayed or difficult to interpret (Levitt and March, 1988; Zollo, 2009). By contrast, we show that decision makers may also learn a different, more detailed representation of decision problems when they are persistently exposed to causal ambiguity and representation learning occurs. Our theory therefore implies a reinterpretation of what causal ambiguity and uninformative feedback mean for a decision maker's ability to learn from experience.

Nonetheless, representation learning will be less applicable to situations where feedback is timely and precise, and trial-and-error changes can be iterated quickly and almost automatically. All else being equal, the possibility to learn from semi-automatic adjustments leads decision-makers to be less likely to engage in more cognitively taxing learning efforts. For instance, by repeatedly executing an operational routine, underperforming actions can be immediately adjusted until performance is satisfactory across all intermediate steps of the procedure, and decision-making becomes automatic or less-mindful (Nelson and Winter, 1982; Levinthal and Rerup, 2006). We would expect representation learning to have a much lower explanatory power than reinforcement in these, and analogous, contexts.

Our second contribution relates to research on deliberate and mindful theories of organizational learning (Weick, Sutcliffe, and Obstfeld, 1999; Zollo and Winter, 2002). Both strands of theory have emerged to challenge and provide alternatives to the automaticity associated with reinforcement learning. Our conceptualization is consistent with these theories and proposes significant extensions. First, contrary to deliberate and mindful learning, representation learning addresses the central question in experiential learning: all else being equal, what predictions can be made regarding the difference in the decision competence of individuals with varying degrees of experience? Our mechanism explains how the accumulation of experience itself drives improved understanding and decision competence.

Second, these theories assume that feedback is available or at least do not single out a scenario where feedback is unobserved. For instance, the learning benefits of knowledge articulation and codification derive from making explicit, valuable implicit knowledge accumulated by reinforcement, which requires prior observations of feedback in the first place (Cangelosi and Dill, 1965; Zollo and Winter, 2002). More generally, the fact that intentional learning investments result in an improved understanding of the causal structure, is a premise on which these theories are built, and which leaves the role of feedback unspecified.

Finally, deliberate and mindful learning theories do not characterize the type of changes in cognition that generate an improved understanding of the causal structure. On the contrary, we provide a rationale for making predictions in terms of the type of changes in representations that occur in representation learning. The key feature of our argument is path-dependency. New finer-grained conceptual distinctions and contingencies are derived from and consistent with previously acquired representations. This pattern implies idiosyncratic learning trajectories that depend on the unique decision experience of each individual. That is, by analogy with a fitness landscape (Levinthal, 1997), equally experienced individuals may develop distinct yet equally accurate representations of the same decision environment. This central aspect of how decision competence is acquired is left unspecified in deliberate and mindful learning theories and is critical for our understanding of how mental representations and decision competence evolve with experience.

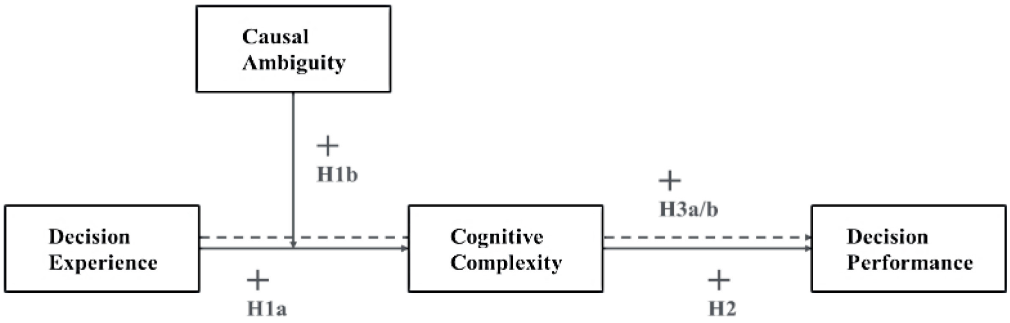
Our third contribution is to the large literature on the role of mental representation in organizational and strategic decision making (Walsh, 1995, Tripsas and Gavetti, 2000). A central question in this literature pertains to the performance implications of the degree of complexity of representations. Less simplified representations that more accurately capture the causal structure of the environment are expected to benefit decision performance (Gary and Wood, 2011). However, managers have limited cognitive capacity for handling high dimensional representations (Levinthal, 2011), and developing highly accurate knowledge is costly at the organizational level (Zollo and Winter, 2002). This tension is reflected by conflicting views of the role of representational complexity (Csaszar and Ostler, 2020), divided among advocates of low complexity and fast-and-frugal heuristics (Gigerenzer and Goldstein, 1996; Sull and Eisenhardt, 2015), of highly complex and accurate representations (Kiesler and Sproull, 1982; Weick, Sutcliffe, and Obstfeld, 1999), and of representations that match the complexity of the environment (Ashby, 1956).

We inform this debate by emphasizing the process by which decision makers develop representations of different complexity. In particular, our main contribution is to show that decision makers may adjust representational complexity not primarily due to its implications for performance, but to adapt to the complexity of the environment and optimize cognitive efficiency. Two implications follow. First, there are often situations where increases in complexity and decision efficiency do not necessarily constitute a trade-off as research on fast-and-frugal heuristics suggests. When decision makers do not already know what the most relevant dimensions of a decision environment are, such as in novel domains or after substantial environmental change, increases in complexity can improve both only efficiency and performance (Csaszar and Ostler, 2020). Put it differently, we suggest that increasing complexity can in fact be a way for decision makers to adapt to environments where learning from feedback is difficult.

Second, we cannot exclude the possibility that decision makers may develop overly complex representations that do not benefit performance when causal ambiguity is especially high and persistent. Research in this stream suggests that there may be diminishing or even negative returns to excessive increases in complexity (Csaszar and Levinthal, 2016; Martignoni, Menon, and Siggelkow, 2016). Excessively complex problem representations may even become intractable from a computational perspective (Bettis and Hu, 2018). We speculate that decision makers may "overfit" representations when they cannot learn from feedback for prolonged periods of time and yet they are induced to keep increasing complexity by persistent perceptions of causal ambiguity. We argued that perceptions of ambiguity diminish as individuals refine their oversimplified representations, but we cannot exclude that conditions may exist under which they remain persistently high – for instance, due to significant and frequent changes in the environment's causal structure. In additional analyses, we obtained preliminary evidence of an inverted curvilinear relationship between decision experience and cognitive complexity. Future research could examine longer decision histories in difficult learning environments and provide more evidence of the relationship between complexity and performance for large stocks of decision experience.

Concluding, in this study, we have developed the concept of representation learning to explain how individuals in organizations can learn, despite an absence or an insufficiency of feedback. For the practice of management, our insights imply that in decision-making contexts where feedback is notoriously delayed or noisy – such as the evaluation of early-stage technologies – organizations are well advised to rely on expert evaluators who are systematically apportioned large numbers of decisions of a given class over time. Additionally, evaluators could be given support for developing requisite complexity of their mental representations and for aggregating representations across the organization. Overall, the key managerial implication is that, even though feedback is not available, it is still possible for an organization to improve its ability to predict outcomes more reliably from decision alternatives as individuals accumulate experience.

Figure 1 – Representation learning, a model of individual learning in organizational contexts.



Dashed lines represent a mediation path.

Table 1 – Examples of how Cognitive Complexity is computed.

Evaluation report text	Rating	Evaluation factors	Causal links	Complexity
*TECH's protocol has STANDARD acceptance, but no application number or official version is available, yet **(1)**. A request tells the user that TECH is done, after which the user is identified. Then the service is identified, and the result can be found in TECH **(2)**. Thus, the STANDARD is covering our patent application **(1)**. The invention is also relevant for different kind of PRODUCTs. Real implementation schedule totally open, though, but possibility exists **(3)**.*	2 => 3	(1) Standardization (2) Technology (3) Implementation	"(2) Thus (1)"	
		3	1	**4**
*The invention is related to TECH used for elimination errors in STANDARD **(1)**. The idea is basically a TECH **(2)** and based on the patent claims **(3)**, the general usability area is left unclear. Some implications are present that it could be used in STANDARD **(1)** but PRODUCT and other TECH techniques might also be possible **(4)** The priority date for the invention is 6.3.1985 **(5)** and the only country where the patent is still active is COUNTRY **(6)**. Thus the technical value of this case is uncertain **(2)** and the patent is anyhow soon to be lapse due the age **(5)**. It is hard to see any business value for this case.*	3 => 1	(1) Standardization (2) Technology (3) Legal claims (4) Implementation (5) Age (6) Geography	"(3) and based on (4)" "(5) Thus (2)" "(6) Thus (2)"	
		6	3	**9**

Terms in upper case represent anonymised references to specific technologies, standards, etc.

Table 2 – Descriptive statistics.

Variable	*Obs*	*Min*	*Max*	*Mean*	*Median*	*SD*
Abandoned Ratio	17,136	0	1	0.11	0	0.18
Age	17,136	0	20	4.52	4	3.40
Backward Citations	17,136	0	249	15.49	12	16.18
Causal Ambiguity	17,136	0.23	1	0.79	0.81	0.12
Causally Ambiguous Patent	17,136	0	1	0.77	0.8	0.17
Cognitive Complexity	17,136	1	12.4	3.25	3.10	0.75
Decision Experience	17,136	1	659	130.52	92	122.29
Decision Performance	17,136	-2	2	-0.32	-1	1.03
Expert Opinion	17,136	0	1	0.08	0	0.28
Family Size	17,136	1	21	4.66	4	3.17
Filing Year	17,136	1990	2007	2001	2002	3.82
Flagged Patent	17,136	0	1	0.36	0	0.48
Forward Citations	17,136	0	525	34.13	20	44.26
Granted Ratio	17,136	0	1	0.17	0	0.27
Inventor Opinion	17,136	0	1	0.14	0	0.34
Number of Claims	17,136	0	120	20.28	20	17.46
Transferred Case	17,136	0	1	0.39	0	0.49
Word Length	17,136	5	789	72.26	37	98.17
Workload	17,136	1	43	4.64	2	7.94

Table 3 – Correlation matrix.

		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)
1	Decision Performance																	
2	Decision Experience	-07																
3	Cognitive Complexity	0.012	0.0562*															
4	Causal Ambiguity	0.0037	-0.2551*	0.0128														
5	Causally Ambiguous Patent	-0.0525	0.2498*	0.0064	-0.5360*													
6	Word Length	-0.0250*	-0.0122	0.1661*	0.0900*	-0.1025*												
7	Inventor Opinion	-0.0336*	-0.0535*	0.0320*	0.1066*	-0.1443*	0.3185*											
8	Expert Opinion	0.0041	0.0126	0.0135	0.0562*	-0.0631*	0.2174*	0.3701*										
9	Granted Ratio	0.0405*	0.1236*	-0.013	-0.4361*	0.7236*	-0.0927*	-0.1242*	-0.0669*									
10	Family Size	0.1215*	-0.0107	-0.0042	-0.0699*	0.1911*	-0.0501*	-0.0418*	-0.0137	0.0943*								
11	Forward Citations	0.0891*	0.0291*	-0.0351*	-0.0066	0.0978*	-0.0476*	-0.0407*	0.0064	0.0799*	0.1653*							
12	Number of Claims	0.0073	0.0376*	0.0217*	0.0891*	-0.1195*	-0.0150*	0.006	0.0142	-0.1390*	0.0542*	0.1412*						
13	Backward Citations	0.0497*	0.0645*	-0.0508*	0.0422*	-0.0111	-0.0311*	-0.0269*	0.014	-0.0523*	0.1620*	0.3343*	0.1186*					
14	Abandoned Ratio	0.0184*	0.1614*	0.0276*	-0.2937*	0.5091*	-0.0330*	-0.0607*	-0.0189*	0.2953*	0.1060*	0.0071	-0.0576*	-0.0207*				
15	Transferred Case	0.0329*	0.1863*	0.0025	-0.1490*	0.3090*	-0.0413*	-0.0118	0.0553*	0.1305*	0.0624*	0.0611*	0.0345*	0.0578*	0.2176*			
16	Workload	-0.0306*	0.0540*	-0.1778*	-0.2945*	0.1338*	-0.1456*	-0.0933*	-0.0622*	0.1729*	0.0220*	-0.0190*	-0.0658*	-0.0283*	0.0739*	-0.0410*		
17	Flagged Patent	0.0956*	0.0018	-0.0139	0.1469*	-0.1030*	-0.0112	0.0439*	0.0674*	-0.1031*	0.1625*	0.0162*	0.0815*	0.0013	-0.0665*	0.1197*	-0.0794*	
18	Filing Year	-0.0614*	0.1222*	0.0106	0.3646*	-0.6350*	0.1330*	0.1231*	0.0981*	-0.5581*	-0.2879*	-0.0579*	0.1947*	0.1027*	-0.2948*	0.0458*	-0.1839*	0.1468*

Observations = 17,136

* p<0.05

Table 4 – Linear regression models of *Cognitive Complexity*.

VARIABLES	**Model 1**	**Model 2**	**Model 3**
Decision Experience		0.16***	0.19***
		(0.01)	(0.01)
Causal Ambiguity			0.01
			(0.01)
Decision Experience # Causal Ambiguity			0.07***
			(0.01)
Word Length	0.00***	0.00***	0.00***
	(0.00)	(0.00)	(0.00)
Causally Ambiguous Patent	-0.8**	0.12	0.17
	(0.08)	(0.11)	(0.11)
Inventor Opinion	-0.05**	-0.05**	-0.05**
	(0.02)	(0.02)	(0.02)
Expert Opinion	-0.01	-0.01	-0.00
	(0.02)	(0.02)	(0.02)
Granted Ratio	0.013	-0.01	0.00
	(0.04)	(0.04)	(0.04)
Family Size	-0.00	-0.00	-0.00
	(0.00)	(0.00)	(0.00)
Forward Citations	0.00*	0.00*	0.00*
	(0.00)	(0.00)	(0.00)
Number of Claims	0.00	0.00	0.00
	(0.00)	(0.00)	(0.00)
Backward Citations	-0.00	-0.00	-0.00
	(0.00)	(0.00)	(0.00)
Abandoned Ratio	-0.00	-0.00	-0.00
	(0.04)	(0.04)	(0.04)
Transferred Case	-0.03**	-0.04**	-0.044***
	(0.02)	(0.02)	(0.02)
Workload	-0.02***	-0.03***	-0.02***
	(0.00)	(0.00)	(0.00)
Flagged Patent	-0.03*	-0.03**	-0.04**
	(0.02)	(0.02)	(0.02)
Constant	2.34***	2.17***	2.13***
	(0.1)	(0.1)	(0.1)
Observations	17,136	17,136	17,136
Number of Patent Engineers	146	146	146
Patent Engineer FEs	Yes	Yes	Yes
Filing Year FEs	Yes	Yes	Yes
Technology FEs	Yes	Yes	Yes

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 5 – Ordered logit models of *Decision Performance.*

VARIABLES	**Model 4**	**Model 5**	**Model 6**	**Model 7**
Decision Experience		0.08**		0.06*
		(0.03)		(0.3)
Cognitive Complexity			0.07***	0.07***
			(0.02)	(0.02)
Word Length	-0.00	-0.00	-0.00	-0.00
	(0.00)	(0.00)	(0.00)	(0.00)
Causally Ambiguous Patent	-0.16	0.27	-.01	0.27
	(0.22)	(0.29)	(0.22)	(0.29)
Inventor Opinion	-0.18***	-0.18***	-0.18***	-0.18***
	(0.05)	(0.05)	(0.05)	(0.05)
Expert Opinion	0.1	0.1	0.1	0.1
	(0.06)	(0.06)	(0.06)	(0.06)
Granted Ratio	-0.00	-0.01	-0.01	-0.01
	(0.01)	(0.1)	(0.1)	(0.1)
Family Size	0.06***	0.06***	0.06***	0.06***
	(0.1)	(0.1)	(0.1)	(0.1)
Forward Citations	0.00***	0.00***	0.00***	0.00***
	(0.00)	(0.00)	(0.00)	(0.00)
Number of Claims	-0.00	-0.00	-0.00	-0.00
	(0.00)	(0.00)	(0.00)	(0.00)
Backward Citations	0.00***	0.00***	0.00***	0.00***
	(0.00)	(0.00)	(0.00)	(0.00)
Abandoned Ratio	-0.11	-0.11	-0.11	-0.11
	(0.11)	(0.11)	(0.11)	(0.11)
Transferred Case	0.06	0.06	0.06	0.06
	(0.04)	(0.04)	(0.04)	(0.04)
Workload	-0.01***	-0.01***	-0.01***	-0.01***
	(0.00)	(0.00)	(0.00)	(0.00)
Flagged Patent	0.27***	0.26***	0.26***	0.26***
	(0.04)	(0.04)	(0.04)	(0.04)
Observations	17,136	17,136	17,136	17,136
Number of Patent Engineers	146	146	146	146
Patent Engineer FEs	Yes	Yes	Yes	Yes
Filing Year FEs	Yes	Yes	Yes	Yes
Technology FEs	Yes	Yes	Yes	Yes

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 6 - Results of bootstrapping moderated mediation GSEM analysis.

	B	Bootstrap SE	*p*-value	95% CI	
INDIRECT EFFECT					
Path *a*					
Experience	0.19	0.01	0.000	0.17	0.21
Experience x Causal Ambiguity	0.07	0.01	0.000	0.06	0.08
Path *b*					
Complexity	0.07	0.02	0.003	0.03	0.13
Path *ab* (CIE)					
CIE – Low Causal Ambiguity	0.01	0.00	0.009	0.00	0.02
CIE – Mean Causal Ambiguity	0.01	0.00	0.005	0.00	0.03
CIE – High Causal Ambiguity	0.02	0.01	0.005	0.01	0.04
DIRECT EFFECT					
Path *c*					
Experience	0.06	0.04	0.123	-0.01	0.13
MODERATED MEDIATION					
$CIE_{mean} - CIE_{low}$	0.005	0.00	0.005	0.00	0.01
$CIE_{high} - CIE_{mean}$	0.005	0.00	0.005	0.00	0.01

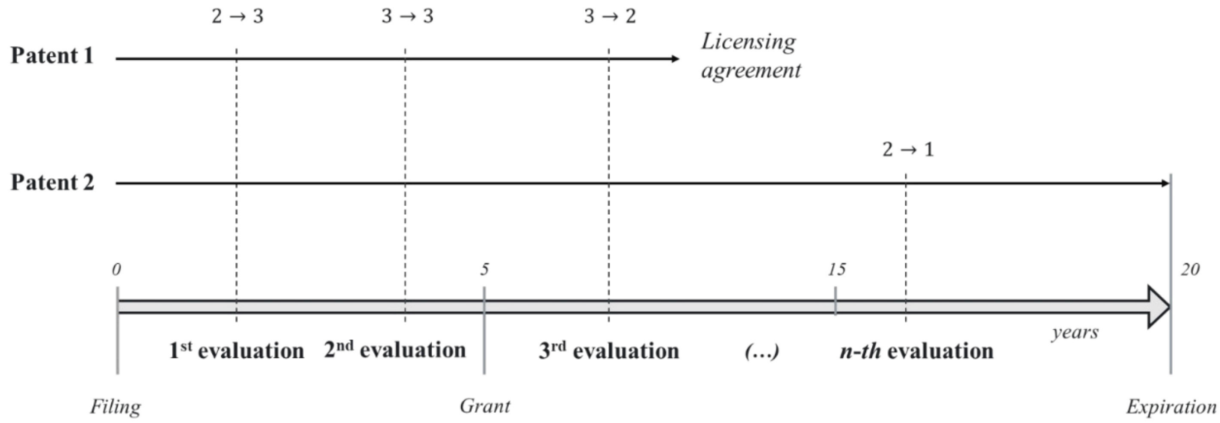
Note: Generalized structural equation modelling (GSEM) fits a single model and estimates both indirect and direct effects (Hayes 2013; Preacher et al., 2007), in contrast to traditional mediation analysis, which involves a series of linear regression models (Baron and Kenny, 1986). A key advantage of this approach is that it allows the residuals to vary (Shaver, 2005). We employed bootstrapping with 1,000 replications and robust standard errors to test the significance of the indirect paths from the independent variable (experience) to the dependent variable (performance) through the mediator (complexity). Generalized SEM allowed us to include patent engineer, filing year, and technology fixed effect, and to use ordered logit models for estimating the dependent variable performance.

APPENDIX A

Table 1 – Definition of Decision performance

Patent engineer's decision (*ex-ante*)		Patent outcome (*ex-post*)	
		A. Positive outcome	*B. Negative outcome*
Rating confirmed or increased	1. *Positive prospect*	**Positive**	Negative (**Type I Error**)
Rating decreased	*2. Negative prospect*	Negative (**Type II Error**)	**Positive**

Figure 1 – Example of how *decision performance* is computed.



		1st evaluation	**2nd evaluation**	**3rd evaluation**	***n-th* evaluation**
Patent Engineer 1	Rating change (ΔR)	1	0	-1	
	Performance	**1**	**1**	**-1**	
Patent Engineer 2	Rating change (ΔR)	1	0	-1	-1
	Performance	**-1**	**-1**	**1**	**1**

In this example, patent engineers 1 and 2 are respectively assigned patent 1 and 2 for periodic evaluations. The timeline shows a sequence of relevant events; for instance, both patents are granted 5 years after filing and the first two evaluations are made while patents are under prosecution. Patent 1 is successfully licensed to a third party after the grant, while patent 2 expires at the end of its legal life without being commercialized. That is, the outcomes for patent 1 and 2 are respectively positive and negative. For illustrative purposes, both patent engineers evaluate their respective patents at the same time and assign the same ratings. Both ratings are increased from 2 to 3 at the first evaluation, confirmed at 3 at the second evaluation, and decreased to 2 at the third evaluation. Patent 1 is removed from the first patent engineer's list of assigned patents because it was commercialized, and additional evaluations are no longer necessary. Patent 2, instead, keeps being evaluated until the expiration date is reached. The table shows that rating changes (ΔR) are equal for both patents, while the sign of decision performance is the opposite. For instance, the rating decrease decided at the third evaluation ($\Delta R = -1$) corresponds to a negative performance for engineer 1 and to a positive performance for engineer 2.

APPENDIX B

Table 1 – Evaluation Factors

	Topic	Factor
1.	Standard	Relevance in Standardization
2.	Implementation	Internal Implementations
3.	Interdependence	Relevance to other Internal Projects and Cases
4.	Business strategy	Relevance to Internal Business Strategy
5.	Competition	Competitors Implementations
7.	Competition	Relevance to Competitors Business Strategy
8.	Commercialization	Relevance to Markets and Users
9.	Commercialization	Relevance to Products and Services
10.	Litigations	Use in Infringements
11.	Litigations	Use in Litigations
12.	Licensing	Relevance for Licensing
13.	Technology	Relevance of the Technology Area
14.	Technology	Control over Technology Area
15.	Technology	Relevance of the Technical Problem
16.	Technology	Comparison of the Invention to other Solutions to the same Problem
17.	Claims	Scope of Protection of Claims
18.	Patent Citations	Patent Forward Citations
19.	Costs	Annuity Costs
20.	Costs	Prosecution Costs
21.	Technology	Possibility to Design Around
22.	Geography	Geography
23.	Claims	Supervision
24.	Claims	Detectability
25	Prosecution	Industrial application
26	Prosecution	Inventive step
27	Prosecution	Prior Art/Novelty
28.	Commercialization	Relevance for Divestment/ Other commercialization efforts

CHAPTER 2

A MODEL OF REPRESENTATION AND POLICY SEARCH IN COMPLEX ENVIRONMENTS

ABSTRACT

Mental representations determine how individuals and firms make decisions and have implications for individual and firm performance. A recent line of work has started investigating the performance implications of searching over alternative representations, including the implications of searching over different representations of dimensions of performance and of various degrees of representational complexity. However, current research does not distinguish between ways of developing representations of different complexity across distinct dimensions of a decision problem. In this study, I explore the trade-offs associated with the allocation of representation search efforts across the distinct dimensions of a decision problem – that is, with the *breadth* of representation search strategies. To this end, I develop a NK model of dual search over policies and representations where agents can either refine their representations broadly across dimensions or deeply in one or few dimensions of a decision problem. Results obtained with this model show that the optimal representation search breadth is contingent on the complexity of the decision environment. Contrary to previous research, intermediate levels of search breadth are associated with optimal performance only for moderate levels of complexity. Higher levels of complexity demand narrow search strategies, while broad search strategies are optimal when complexity is low. A second set of results explores the relationship between the breadth of search strategies and the optimal degree of representational complexity. In line with recent findings in this research stream, I find that, counterintuitively, less accurate representations can outperform more accurate ones – i.e., that the optimal degree of representational complexity does not necessarily match the true complexity of the environment. However, I show that less accurate representations can outperform more accurate ones only for broad rather than narrow representation search strategies. These findings contribute to research on learning and adaptation in complex environments and on the role of mental representation in organisational decisions.

INTRODUCTION

Mental representations determine how individuals and firms make decisions (Gardenfors, 2004; Simon 1990) and have implications for individual and firm performance (Barr, Stimpert, and Huff, 1992; Eggers and Kaplan, 2009). Representations vary across individuals and over time, for instance, in terms of their accuracy or degree to which they simplify distinctions among the features of the environment (Gary and Wood, 2011; Axelrod, 2015; Martignoni, Menon, and Siggelkow, 2016). Changes in representations occur with experience (Walsh, 1995) and as the result of search efforts (Csaszar and Levinthal, 2016), thus their evolution is core for understanding the origins of individual capabilities and their implications for firm performance.

A recent line of work has started investigating the performance implications of searching over alternative representations (Csaszar and Levinthal, 2016; Martignoni, Menon, and Siggelkow, 2016; Choi and Levinthal, 2022). Managers can and do update their representations of problems to be solved and of the decision environment as they attempt to achieve superior decision performance (Fiske and Taylor, 1984; Barr, Stimpert, and Huff, 1992; Benner and Tripsas, 2012). Recent work has started addressing fundamental questions about the implications of searching over different representations of dimensions of performance (Csaszar and Levinthal, 2016), of various degrees of representational complexity (Martignoni, Menon, and Siggelkow, 2016; Csaszar and Ostler, 2020), and of different ways of encoding prior experiences (Choi and Levinthal, 2022).

While representational complexity is central to our understanding of the performance implications of representations, current research does not distinguish between ways of developing representations of different complexity and level of detail across distinct dimensions of a decision problem. This distinction is important for two reasons. First, different ways of developing complexity and refining one's understanding of decision problems reflect a strategic trade-off in how representation search efforts are allocated across dimensions. Managers can, on the one hand, focus their search efforts to find increasingly accurate representations of only one or few aspects of a decision problem; or, on the other, allocate search efforts more broadly and look for new ways of representing problems from multiple perspectives. Thus, for any given level of experience and search effort, individuals can only develop a more granular yet narrower understanding of a problem in the former case, or a broader yet less accurate understanding in the latter (Teodoridis, Bikard, and Vakili, 2019).

Second, different representations search strategies across dimensions of a decision problem have important performance implications. On the one hand, using increasingly accurate and finer-grained representations of one or a few aspects of a decision problem can have large performance implications because decision environments are complex and control over even small details can have system-wide effects. To illustrate, consider the case of the two de Havilland Comet jets that exploded mid-air on their way from and to London in 1954. The two fatal plane crashes were caused by the design of the planes' windows, which were square at the time. Investigations after the incidents confirmed that the sharp corners of the windows put the surrounding metal under up to three times the stress endured by other parts of the plane, resulting in the destructive cracks that eventually caused the systemic failures. Engineers likely held a sophisticated understanding of the window structure design problem, given how critical structural integrity is in aeronautical applications. Further, the intensification of stress around sharp corners was a known phenomenon at the time. It is reasonable to assume that designers could have anticipated the importance of avoiding sharp corners had they allocated efforts to further refine their representation of the design problem.

On the other hand, narrow and focused search efforts may have limited benefits when other aspects of a problem are oversimplified. Interaction effects among the dimensions of a decision problem imply that the improvements obtainable by refining a subset of choices may largely depend on other underlying distinctions that managers neglect. For instance, innovative products that offer sophisticated new features or market strategies based on differentiation may fail because of critical shortcomings with other underdeveloped features (Barwise and Meehan 2004). The history of the Ford Pinto is not only a case of questionable business ethics but also of basic design flaws that could have been easily avoided. Engineers realised during final tests before the product launch that even low-speed rear impacts could cause fires because the fuel tank was simply too close to the rear bumper. The Pinto was praised at the time for being the first American car in the small-car segment to feature advanced solutions for low weight, low costs and a still competitive delivery deadline. Nonetheless, the original design evidently overlooked critical interactions between choices on a larger scale, such as the ones between the positions of the rear bumper and fuel tank. From this perspective, the representation of the design problem would have benefited from a broader allocation of search efforts across other decision variables.

In this study, I explore the trade-offs associated with the allocation of representation search efforts across the distinct dimensions of a decision problem – that is, with the *breadth* of representation search strategies. To this end, I develop a NK model of dual search over policies and representations (Csaszar and Levinthal, 2016). Search over policies occurs as in traditional NK models of search where agents iteratively try choice alternatives across distinct categories of policy choices[13] (Gavetti and Levinthal, 2000). In terms of search over representations, I focus on search efforts by which agents attempt to refine their representations of each policy category. In this sense, changes in representation correspond to situations where individuals realise that choices that they had treated as equivalent have instead distinctive features that may lead to distinct outcomes (Choi and Levinthal, 2022). For instance, a refinement of the aeroplane's window design problem could result in disaggregating the choice of the outer shape into two separate choices, e.g., of a width-to-height ratio and of whether shapes present sharp corners. Accordingly, the breadth of a representation search strategy is related to the number of distinct categories of policy choices across which decision-makers allocate representation search efforts.

I incorporate search over representations by adding two components to the traditional search model. First, the model features two search landscapes, namely the true performance landscape (of dimension N) and agents' simplified representation of it (of dimension $N_a \leq N$). Each of the N_a agents' policy category aggregates a distinct set of the N dimensions of the environment, such that more distinctive categories aggregate fewer dimensions and are more accurate. Agents search over policies according to their simplified representations of possible choice alternatives, while payoffs are computed according to the true landscape. Second, agents can refine their representations by replacing one coarser policy category with two more distinctive ones. As such, representations become increasingly more accurate (N_a approaches N) as agents search over representations. The set of categories that agents can refine

[13] Traditional NK models of search define N categories of policy choices and two choice alternatives (or 0/1 "bits") for each policy category.

depends on their representation search breadth, such that broader strategies include a larger number or all the (N_a) policy categories.

Results obtained with this model show that the optimal representation search breadth is contingent on the complexity of the decision environment. Contrary to previous research, intermediate levels of search breadth are associated with optimal performance only for moderate levels of complexity. Higher levels of complexity demand narrow search strategies, while broad search strategies are optimal when complexity is low. These results can be explained by noting that simplified mental representations generate *apparent* local peaks of performance that do not correspond to *true* local peaks of the performance landscape. That is, decision-makers stop searching for new policies when they believe they have exhausted all local possibilities for improvement. However, the mapping between their simplified representations of available policy choices and the true performance landscape indicates that this often happens when they are not located at a local peak of the true landscape and that further performance improvements are in fact available even locally. Put it differently, choices taken from the perspective of a simplified representation can have consequences on the true performance landscape that decision makers simplify and neglect but that can have significant implications for performance.

Importantly, the simple model of this study also shows that feedback noise is not exclusively a property of the external environment but also a consequence of representing choice alternatives by means of simplified representations. A key implication of simplification is that what decision makers treat as one policy choice in fact affects multiple features or dimensions of the environment in a way that they do not observe or are unaware of. It follows that when a policy is changed from the current state to a new alternative, the actual implementation of the alternative affects the environments in ways that decision makers do not control and that are thus random. The feedback signals they observe are thus noisy – i.e., configurations of policy choices generate payoffs that are non-deterministic – even when the payoffs of the true performance landscape are deterministic.

Finally, a second set of results explores the relationship between the breadth of search strategies and the optimal degree of representational complexity. In line with recent findings in this research stream, I find that, counterintuitively, less accurate representations can outperform more accurate ones – i.e., that the optimal degree of representational complexity does not necessarily match the true complexity of the environment (Csaszar and Levinthal, 2016). However, I show that less accurate representations can outperform more accurate ones only for broad rather than narrow representation search strategies.

This study contributes to the literatures on search and the performance implications of mental representation. I introduce the notion of apparent performance peak as a necessary consequence of simplified mental representations that affects how decision-makers conduct policy search. Contrary to the assumptions of traditional local policy search, this notion suggests that decision-makers may stop searching at locations of the landscape that do not correspond to local peaks and that are thus suboptimal. In other words, decision-makers may often stop searching when they are one small policy change or “step” away from a location associated with higher performance. From a managerial perspective, these results suggest that managers do not necessarily need to explore distant policy configurations when they believe they do not have other possibilities for improvement locally. Instead, managers can just refine their representations of available policy choices and find opportunities for improvement by modifying their configurations locally at finer level of detail.

Finally, this study contributes to the literature on the trade-offs between specialist and generalist knowledge. By leveraging the notion of representation search, I identify conditions under which it is worth investing in narrowly and deeply refined “specialised” representations rather than in broadly and uniformly developed “general” representations. Research on organisational search suggests that optima are found at intermediate levels of search breadth (Katila and Ahuja, 2002), while research on knowledge specialisation and creativity has produced inconsistent findings and found support for either high or low levels of specialisation (Teodoridis, Bikard, and Vakili, 2019). This study contributes to these literatures and shows that the level of interdependence among policy choices is a critical contingency affecting the performance of different levels of knowledge specialisation.

THEORETICAL BACKGROUND

Mental representations are central to decision-making because the set of choice alternatives that decision makers can control and the performance outcomes they can observe and learn from depend primarily on how they represent decision environments in the first place. Decision-makers can process only a portion of the information available for a decision task (Simon, 1995). This insight has been foundational for the behavioural theory of the firm (March and Simon, 1958; Cyert and March, 1963; Gavetti et al., 2012) and the behavioural economics (Kahneman and Tversky, 1979) research traditions, which maintain that decision-makers most often rely on simplified representations of a decision environment. Using a simplified mental representation for making decisions entails neglecting distinctions among entities and interdependencies of the environment (Simon, 1990). It follows that the consequences of choosing among a set of choice alternatives depend primarily on the extent to which alternatives capture distinctions among environmental features, because decision-makers cannot observe or have control over features that their representations neglect (Winter, 1987). For instance, a very simplistic representation of an aeroplane design task may involve the choice of only the wing total surface area and of the fuselage length. This two-dimensional representation is over-simplified because a designer would not have control over other critical sets of choices such as engine features or the internal structure of wings and fuselage.

Research shows that decision-makers do indeed search and learn different representations of decision tasks and that changes in representation are highly consequential for performance (Fiske and Taylor, 1984; Walsh, 1995; Tripsas and Gavetti, 2000). Since representations differ primarily in terms of which environmental features are captured or neglected, representation search entails changes in how the choice alternatives available to decision-makers are represented. Each set of choice alternatives represents an aspect or dimension of a decision task. Alternatives within a set are distinct because they differ in at least one environmental feature that characterises the respective dimension. The fewer features across which alternatives differ, that is, the larger the number of entities and interdependencies that are neglected, the more simplified and less accurate representations are (Rosch et al., 1978; Walsh, 1995). Searching for representations involves deciding which of the known environmental features are considered with the sets of choice alternatives available to decision-makers. For instance, with reference to one of the opening examples, the presence of sharp corners in the outer shape of aeroplanes' windows was not a feature represented by an independent set of choice alternatives in earlier representations of the aeroplane design task and designers did not have direct control over it. As this example suggests, the consequences of failing to change current representations can be large. Similarly, research shows that survival after large environmental shifts requires searching for different representations of key decision variables and interdependencies (Barr et al. 1992; Tripsas and Gavetti, 2000).

It follows that while superior decision performance depends on choosing the optimal configuration of policy choices, the possibility itself of choosing the optimal configuration depends on how the sets of choice alternatives are represented in the first place (Gavetti and Levinthal, 2000). Thus, I refer to the *optimal representation* for a given decision environment as the set of representations that maximise the performance attainable via local policy search (i.e., via "hill-climbing").

Two considerations are critical to understand the search for optimal representations. First, highly accurate representations are not necessarily optimal. Different literatures present conflicting views of the performance consequences of representational accuracy (Csaszar and Ostler, 2020). Research has examined the performance advantages of simple and fast-and-frugal heuristics (Gigerenzer and Goldstein, 1996; Sull and Eisenhardt, 2015), of highly complex and accurate representations (Kiesler and Sproull, 1982; Weick, Sutcliffe, and Obstfeld, 1999), and of representations that match the complexity of the environment (Ashby, 1956). Advocates of simple representations suggest that simplified decision rules or heuristics can be applied across contexts and free up resources to address the specific features of different environments (Davis et al., 2009). Heuristics may even be a natural evolutionary response to deal with uncertainty, often superior to human attempts at rational optimisation (Gigerenzer and Gaissmaier, 2011). A recent stream of simulation studies has advanced this debate by examining contingencies that influence the optimal degree of representational accuracy. While greater accuracy is often advantageous (Simon, 1990), excessively accurate representations may be detrimental to performance depending on characteristics of the environment, such as the degree of interdependence among policy choices (Martignoni et al. 2016; Csaszar and Levinthal, 2016); and on characteristics of the decision maker, such as levels of experience and of knowledge of critical dimensions of the decision task (Csaszar and Ostler, 2020).

Second, even assuming that any representation can be potentially learned given sufficient time and resources, not all representations can be adopted within any given time horizon and decision-makers must decide how search efforts are allocated across different dimensions of representation. It is reasonable to assume that representation search occurs locally and gradually akin to traditional models of policy search (Cyert and March, 1963). Decision-makers tend to avoid distant search because changing many policy variables at once is risky (Levinthal and March, 1993). By the same token, decision-makers likely avoid changing large portions of current problem representations at once (Csaszar and Levinthal, 2016). It follows that similar to traditional models of experiential learning (Gavetti and Levinthal, 2000), mental representation learning occurs gradually and in a path-dependent fashion. Earlier choices as to which dimensions should be searched for improved representations would make some subsequent representation changes possible while making others unattainable, especially in the presence of search interdependencies and under the constraint of limited time and resources (Levinthal, 2021). Hence, the possibility of attaining optimal representations depends on how decision-makers allocate representation search efforts across different dimensions of a decision task.

In this paper, I focus on the latter aspect of representation search and examine the performance implications of different ways of allocating search efforts across distinct dimensions of a decision problem. While the implications of any given degree of representational accuracy have received most of the research attention in this stream, it is equally important to examine the consequences of different search strategies, that is, of different ways of searching for optimal representations. Answering this question is important because it may reveal, for instance, that representations are not equifinal – that is, that the likelihood of attaining an optimal configuration of policy choices via local policy search depends on the representation search trajectory followed by a learning agent rather than on the representation per se, and thus on how search efforts are allocated.

Specifically, I focus on the implications of searching broadly rather than narrowly across representational dimensions. Indeed, this distinction reflects the traditional notion of (policy) search breadth (Katila and Ahuja 2002). In the context of representation search, breadth is defined as the extent to which decision-makers adopt changes in representations across a broad rather than narrow set of dimensions of a decision problem. For instance, a phone manufacturer may adopt relatively simple sets of choice alternatives for the operating system, design, and battery features; and devote their limited search resources to change, and ideally improve, the representation of camera and display features. The manufacturer may experiment with different choice alternatives, that is, search over policies across all the dimensions of the phone design. However, the choice of narrowly focusing representation search efforts on two dimensions reflects the manufacturer's strategic interest in maximising the chances of attaining a high-performance configuration of policy choices for these two sets of phone design features.

In order to emphasise the distinction between broad and narrow search strategies, I examine the representation search process by making two key simplifying assumptions. First, decision-makers use mental representations that are initially over-simplified with respect to the distinct features of a decision environment. Second, they search representations by adopting increasingly accurate sets of choice alternatives along the dimensions specified by their chosen search breadth strategies. Indeed, this setup is simplified in many ways. For example, representations can be over-specified rather than excessively simple (Martignoni et al. 2016); that is, they can account for distinctions among features that are irrelevant for performance or even spurious, i.e., superstitious (Levitt and March, 1988; Zollo, 2009). In this case, representation search may operate in the opposite direction by discounting distinctions among features and aggregating sets of choice alternatives (Choi and Levinthal, 2022). More broadly, representation search may entail changes not only in the degree to which dimensions accurately capture distinct entities and interdependencies but also in how interdependencies among choices are represented (Barr et al., 1992) or in the understanding of how representational dimensions affect the overall performance (Csaszar and Levinthal, 2016; Csaszar and Ostler, 2020).

However, this simplified model retains the essential traits of representation search in a parsimonious fashion while giving salience to the central distinction between broad and narrow search strategies. In particular, overly simplified initial representations imply that decision-makers very likely benefit from increasing representational accuracy (Simon, 1990; Csaszar and Ostler, 2020). While the second assumption ensures that accuracy increases because of any search effort across any of the representational dimensions. Thus, all else being equal, any difference in the efficacy of representation search is due to differences in the chosen representation search breadth strategy. In other words, this model can answer questions about whether focusing limited search resources narrowly and deeply on a limited set of dimensions has different performance implications than allocating resources more broadly and uniformly across several or all the representational dimensions.

In the following sections, I further specify the model of this study and derive a set of results concerning the implications of different representation search strategies.

A MODEL OF MENTAL REPRESENTATION AND SEARCH

The strategy context and mental representation

There are multiple and equally valid ways to incorporate mental representation into models of organisational search (Csaszar and Levinthal, 2016). The richness of perspectives and theories developed, for instance, in artificial intelligence or the cognitive sciences, offers a variety of conceptual tools that can be useful in strategy-making contexts. A parsimonious choice of tools thus depends on the model's objectives (Adner et al., 2009). In order to justify a choice, I first clarify what the strategy-making context involves in this study and then describe a minimal set of elements that can be included in a model of mental representation and search.

For the purposes of this study, the strategy context only needs to comprise two of three elements described by Adner, Csaszar and Zemksy (2014), namely policies and firm profits. Policies are a collection of decisions and actions that managers can take to control their environment, and firm profits measure the market response to the outcomes of policies [14]. For instance, a car manufacturer can control the type of engine or the number of seats of a car (policies), which jointly determine speed and market category (performance dimensions). In turn, speed and market category most directly influence the market response (firm profits). Equivalently, organisational design choices about the type of supervision at the manufacturer's production floor (policies) have consequences for employee retention and production quality (performance dimensions) that ultimately affect the market response (firm profits). With respect to traditional NK models, managers search over N categories of policy choices, such as "engine type" or "supervision policy", to find a configuration that maximises fitness with the landscape, that is, profits.

[14] The third element, performance dimensions, logically links policies to profits: it refers to the outcomes of policy choices, such as product functionalities or the consequences of strategies, that determine the market response.

Policies and profits can be linked to mental representation by means of two sets of elements, namely cognitive categories and payoff expectations. Individuals encode policies according to the conceptual categorisation systems of their mental representations and make policy choices depending on their expected payoffs. Mental representations are often conceptualised as cognitive categorisation systems in which categories represent sets of cognitive objects that individuals perceive as sufficiently similar (Rosch, 1978; Murphy, 2002; Thagard, 2005). Cognitive objects are choices and actions in decision-making contexts, and similarity may depend on whether choices are perceived as being causally linked to the same outcomes or goals (Rehder, 2003; Barsalou, 1991). When managers make a design choice or implement a policy, they select among choice and action alternatives that they believe will achieve the same goal, although differently and with different payoffs. Thus, there is a one-to-one correspondence between policies and the cognitive categories of agents' mental representations. For instance, "combustion" or "electric" are two policy choice alternatives associated with the category "engine type".

In turn, managers choose an alternative depending on policies' expected payoffs. Decisions are said to reflect forward-looking or "theory-driven" forms of intelligence in these cases (Felin and Zenger, 2017). In the strategy context, managers make policy choices guided by their expected consequences for firm profits. For instance, the choice between two rather than five car seats may depend on designers' expectations about the potential profits associated with the respective market segments. Together, categories of policies and expected payoffs allow us to analyse the strategy context through the lens of mental representation.

Dual search over policies and representations

Managers can solve problems by trying policy choice alternatives or by finding new ways to represent policy choices and the problems they are meant to solve. Regardless of whether decisions are guided by rational calculations of expected payoffs or other mechanisms, the choice of a policy alternative inevitably depends on how policy choices and the strategy context are represented (Levinthal, 2011). In this study, managers can search over alternative representations that differ by the degree of accuracy with which they account for distinctions among policy choice alternatives. Coarse representations use simple and inclusive categories according to which many distinct choice alternatives are perceived as equivalent and are associated with the same policy choice (Choi and Levinthal, 2022, Martignoni et al., 2016). Managers can search over representations by refining coarse categories and accounting for finer-grained distinctions among alternatives. A coarse category is refined by disaggregating one or more of the distinct alternatives it represents and associating them to a different, additional category. The additional category thus becomes one of the policy choices that managers can consciously and deliberately control. In other words, mental representations are refined by replacing a coarser category with two more distinctive categories of policy choices. For instance, the choice of a car engine can be refined by choosing not only between combustion and electric but also the maximum output power it can generate.

The two central elements of this model are (i) a mapping between categories of policies and the individual alternatives they represent and (ii) the type of search over policies that this mapping induces. I refer to the choice alternatives of the *true* landscape as the variables of the strategy context at the highest degree of accuracy or granularity that agents could potentially control [15]. Categories of policies aggregate multiple true choice alternatives such that each alternative is associated with one category, but one category may represent multiple alternatives. That is, there is a one-to-many mapping between the N_p policy categories of agents' representations and the $N > N_p$ true choice alternatives. A key feature of mental representation is that agents are either unaware of distinctions among aggregated choice alternatives or believe that the alternatives associated with the same category result in similar outcomes with similar payoffs. In other words, each category of policies corresponds to a decision variable that agents believe can control.

From this mapping follows that when agents search over policies, the changes they make to one of the N_p policy categories can influence any of the aggregated true choice alternatives associated with them. This occurs because agents interpret and evaluate the difference between their initial configuration and the new configuration they intend to try from the perspective of their simplified representations, which do not account for the new state of the respective true choice alternatives. To illustrate, assume that the

[15] True choice alternatives can be thought of as being determined by available technology or knowledge.

choice of a car engine can be simplified to such an extent as to represent only the difference between combustion and electric engines. A car manufacturer produces combustion engine cars and wants to try an electric engine version. From the perspective of this simplified representation, this policy change can be implemented by choosing *any* electric, "non-combustion" engine, regardless of other features. In this hypothetical scenario, the manufacturer has no control over other finer-grained choice alternatives or variables related to the new engine type, such as weight or size. Nonetheless, the new design will have to go into production and eventually to the market. The choice of the actual engines mounted on the new car models will have to be made by others down the production line, such as by engineers at the production plant or by the car suppliers. In this way, all the engines will be electric but will likely present different finer-grained features. The car manufacturer will likely observe variations in the market responses as a consequence of these differences, for instance, in different profits across regions.

Accordingly, I assume that agents search over policies by changing the state of one of the N_p policies of their representations and that payoffs are computed with respect to changes in a random number of true choice alternatives associated with the chosen policy category. For instance, an NK performance landscape may be defined by $N = 16$ true choice alternatives. A simplified representation initially accounts for $N_p = 4$ categories of policies, in which each category aggregates four true alternatives. Agents' *perceived* configuration is defined by the state of the four policy categories, while their *true* configuration on the landscape is defined by the state of the 16 true choice alternatives. When search over policies occurs, agents change the state of one of the four policy categories. The new true configuration is determined by changing the state of a random number of true alternatives among the four alternatives associated with the chosen policy. Agents observe the payoff of the new configurations and decide whether to retain or revert the policy change depending on whether performance has improved. If the change is reverted, both the perceived and true configuration returns to their previous states.

Search over representations occurs by increasing the number N_p of policy categories and decreasing the number of true choice alternatives associated with them. One category is chosen and replaced with two more distinctive categories that each randomly aggregate a subset of the true choice alternatives of the replaced category. With reference to the previous example, assume that the choice of a car engine aggregates the choice of an engine type with three other features, such as the maximum power output, size, and weight. Manufacturers can refine their representation by replacing this coarse "engine" category with two more distinctive categories that aggregate, for instance, engine type and size in one category and power output and weight in the other. This example is unrealistic, but the two examples of the Comet crashes and the Ford Pinto could be described in similar terms. True choice alternatives, such as the presence of sharp corners or the distance from the rear bumper, were perceived as inconsequential and neglected until representations were refined. Eventually, these distinctions became part of designers' deliberate choices.

Representation search breadth and agents' search heuristic

The objective of this study is to determine the relative advantages of broad and narrow representation search strategies. Thus, I focus on identifying the optimal balance between incrementally refining one policy category on the one hand and spreading search efforts across all N_p representational categories on the other. Representation search breadth increases with the number b of categories of agents' initial representations that are incrementally refined.

I model agents' search process as follows. At each time step, one representation search iteration precedes one policy search iteration. At the beginning of each time step, one policy category is selected among the b categories of the initial mental representation defined by agents' search breadth strategies. Either the selected category was never refined before, or it has already been refined a certain number of times. In the former case, the chosen category is replaced by two more distinctive categories that each aggregate one of the two random subsets of true choice alternatives of the chosen category. For instance, an agent chooses a policy category A that has never been refined before, and that aggregates five true alternatives and replaces it with policy categories A_1 and A_2 that each aggregate 2 and 3 true alternatives, respectively. Instead, if the selected category has already been refined, agents know which of the more distinctive categories have been obtained by refining the selected category. One of the more distinctive categories is randomly chosen and further refined as before. For instance, if policy category A is chosen again, one among A_1 and A_2 is picked at random and replaced with two more distinctive categories. If category A_1 was chosen, it would be replaced by A_{11} and A_{12}, while if category A_2 was chosen, it would be replaced by A_{21} and A_{22}. In this example, A_{11} and A_{12} correspond to the highest degree of accuracy and cannot be further refined. The states of the true choice alternatives (the values of the 0/1 "bits") do not change either in the former or in the latter case; that is, agents do not "move" over the landscape as a result of representation search.

After the representation search iteration, search over policies occurs as usual and concludes the time step. As anticipated above, policy search occurs by randomly selecting and changing the state of one of the available policy categories, including the ones that have just been refined. Payoffs are computed by changing the state of a random number of true choice alternatives associated with the selected category. The change is retained if performance improves; otherwise, both the perceived and true configurations remain the same as they were at the end of the representation search iteration.

Figures 1 and 2 show two examples of narrow and broad representation search strategies. Figure 1 shows that after five representation search iterations, one policy category is highly accurate (A) and the other two are coarse (B and C) as a result of a narrow search strategy ($b = 1$).

<<Insert Figure 1 about here>>

On the other hand, Figure 2 shows that after the same number of representation search iterations, all categories are partially refined as a result of a broad search strategy ($b = 3$), although none of them is as accurate as policy category A in Figure 1.

<<Insert Figure 2 about here>>

To best capture the implications of searching narrowly rather than broadly, representation search stops when the narrowest search strategy ($b = 1$) reaches the highest level of accuracy. In fact, this modelling choice reflects the fact that decision-makers need to allocate limited representation search resources either along one or multiple dimensions.

RESULTS

I set the number of true choice alternatives to $N = 160$[16]. Agents' initial mental representations use $N_P = 10$ policy categories; thus, each policy category initially aggregates sixteen true choice alternatives. If not mentioned otherwise, I report average performance results for simulations of $t = 200$ periods based on 200,000 replications. To provide meaningful benchmarks against which I can evaluate each representation search breadth strategy, I include two benchmark agents who do not search for representations and thus perform only policy search with their fixed initial representations. The first benchmark agent is identified as "*Simple*" and uses a fixed representation with $N_P = 10$ policy categories throughout the entire simulation. The second benchmark agent is identified as "*Fully Accurate*" and uses a fixed representation with $N_P = 160 = N$ policy categories throughout the entire simulation – that is, the latter agent uses a policy category for each of the true choice alternatives. The other agents search for representations with three different search breadth strategies: a first agent searches narrowly for more accurate representations along only one dimension ($b = 1$), a second agent searches broadly along all the dimension ($b = 10$), and a third agent searches along an intermediate number of dimensions ($b = 5$).

Benchmark results: apparent local peaks of performance

In Figure 3, I report the baseline results for the five agents defined above. The results show the performance level on the landscape on the y-axis and the number of simulation time periods on the x-axis. As usual, performance levels are normalised to 1 therefore all agents' initial performance is 0.5. The five panels from A to E indicate five different values of the complexity K, from relatively low complexity ($K = 15$) to high complexity ($K = 145$).

The first result to note is the performance difference between the two benchmark agents who search the landscape for the optimal policy configuration but do not search for representations. There is a

[16] Since representation search stops when the narrowest breadth strategy (b=1) achieves the highest degree of accuracy, this relatively large value for N allows me to simulate a sufficiently large number of representation search steps. For instance, by choosing $N_P = 10$, simulations can compute 15 representation search steps. I developed an algorithm for solving search over NK landscape "on-the-fly" that can process values of N and K larger than 1,000 quickly and efficiently.

substantial performance advantage in using a fully accurate representation (marked red line) as opposed to an overly simple representation (marked green line), although the relative advantage decreases for higher levels of complexity. Note that in the figure, the performance of the fully accurate agent never reaches steady state within the chosen simulation time.

<<Insert Figure 3 about here>>

This large performance difference is due to *apparent* local performance peaks that affect agents who do not use fully accurate representations. As discussed above, policy search along a policy category that aggregates multiple true choice alternatives entails switching the state of a random number of those true alternatives. This is equivalent to distant search in traditional NK models, which is known as a suboptimal search strategy per se because it is more likely to diminish rather than improve performance immediately after a distant search step is performed. That is, when multiple true choice alternatives are switched at random, performance is more likely to be inferior and thus agents are more likely to revert it. It follows that an agent who uses simplified policy categories is more likely to exhaust all the perceived choice alternatives as compared to more accurate representations and she will reach steady-state sooner, as evident in Figure 1. Put it differently, agents believe to have reached a local peak of the performance landscape when their configuration on the true performance landscape does not necessarily correspond to a true local perspective peak. The more a representation is simplified, that is, the lower the number of perceived dimensions N_P, the sooner an agent will stop policy search at an apparent local performance peak.

To illustrate, consider again a car manufacturer whose perceived representation of a car design only (and unrealistically) accounts for three dimensions, namely engine type, number of seats and quality of interiors. As in standard NK models, each policy category has two states or choices, for instance combustion versus electric engine, two versus five seats and economy versus luxury interiors, for a total of $2^3 = 8$ policy configurations. The car manufacturer wants to innovate her traditional combustion engine, five-seater, economy interiors model with known profits and begins by launching an equivalent model with electric engine. The new car model is successful, profits are higher than for the traditional model and she decides to keep searching for an improved configuration of choices. Thus she experiments with a two-seater version of the car, but profits are lower than the five-seater model and this configuration is abandoned. If switching car interiors from economy to luxury also does not improve profits, the car manufacturer will believe that she has exhausted her choices and that she has found a performance peak on the landscape.

However, the manufacturer's representation neglects the full range of implications associated with two rather than five seats. For instance, a distinct set of features such as the geometry of the internal spaces or the presence of back seats (some two-seater cars still have small back seats) are all affected by the choice between two and five seats but are not under the manufacturer's control according to her very simple representation. These finer-grained choices will be made by someone else down the design and production lines. Yet, the car manufacturer will consider all two-seater versions as equivalent regardless of finer-grained choices and simply compare the profits of all the two-seater sales against all the five-seater sales when deciding whether to retain the two-seater configuration[17].

The difference between the Simple and Fully Accurate agents shows that there is an advantage in improving the accuracy of representations under these conditions. Agents who use representations that have more control on true choice alternatives will less likely encounter apparent local peaks and will search for longer, increasing the likelihood of approaching a peak of the true performance landscape. In the next sub-sections, I examine whether differences in the ways more accurate representations are searched affect the likelihood of encountering apparent local peaks.

The breadth of representation search strategies

With reference to Figure 3, I examine the performance over time of narrow (green line), intermediate (blue line) and broad (red line) representation search strategies. The first result is that the complexity of the environment K has a large impact on the efficacy of search breadth strategies. Search breadth has a marked effect on performance for low to intermediate levels of complexity, although it vanishes at very high levels of complexity.

[17] Indeed, this line of reasoning must be interpreted from the perspective of the usual assumptions of NK models and are unrealistic in real-world settings. Nonetheless, the intuition behind apparent local peaks is captured even by this simplified scenario.

Second, broad search benefits performance at low complexity, narrow search performs best between intermediate and higher levels of complexity (up to $K = 110$, Panel D), and intermediate breadth search strategies are optimal between low and intermediate values. This result can be explained by the fact that spreading representation search efforts across multiple dimensions produces sets of policy categories that are still relatively too aggregated – i.e. simple – when complexity is higher. Higher complexity implies that performance is more likely to be inferior when multiple true choice alternatives are switched simultaneously, that is, when random distant search is performed. Hence, it is beneficial to focus representation search on fewer dimensions and have more accurate control over the respective subset of true choice alternatives rather than having lower increases of accuracy across more dimensions.

This rationale also supports the result that search breadth has a relatively low impact on performance in highly complex environments. Having accurate control over the true choice alternatives of one or a few dimensions is not sufficient to generate relative performance advantages because the performance contributions of those alternatives still depend strongly on the state of the true alternatives over which agents do not have control. In other words, it does not matter whether agents can control a subset of finer-grained features because they depend on other finer-grained features that agents do not control and that are changed at random.

As these results show, it is critical to understand the role that the mapping between perceived and true policy choice alternatives has on the efficacy of representation search breadth strategies. The mapping reflects the essence of mental representation in terms of neglecting the distinctions between true choice alternatives. Given that decision-makers do not account for these distinctions, it is reasonable to assume that when they switch one of their perceived policy categories, they affect a random number of the true choice alternative that the selected perceived policy choice aggregates. Accordingly, the more a perceived policy is aggregated and simple, the higher the average number of true choice alternatives that are switched at each policy search step and the lower the likelihood that performance can be improved. It follows that all else being equal, search strategies will have a differential impact on performance if agents could control not only which true choice alternative is switched – which depends on representational accuracy – but also the number of true choice alternatives that are switched within a selected policy category.

I examine the implications of a range of values for this parameter of the model in the next section.

Simplified policy search representations and feedback noise

An implication of making choices by means of simplified representations is that choices' feedback signals can be noisy – i.e., non-deterministic – even when the true payoffs generated by the environment for any given choice of policy choices is deterministic. The reason is that decision makers neglect and do not have control over finer-grained features of the environment when they implement the policies that they choose by means of their simplified representations. Implementation thus affects a random number of the neglected features in an uncontrolled way that can have tangible implications for performance. From the perspective of search models, the traditional NK model assumes without loss of generality that each dimension or policy category has two states or choices (or 0/1 "bits"). Yet, even in keeping with this assumption, an agent who repeats twice the same move from the first state "0" to the second state "1" may end up in different locations of the landscape even if representation search is not performed and other policy choices are not changed. As mentioned above, this occurs because, all else being equal, switching the state of the chosen perceived policy from 0 to 1 changes the state of a random number of true choice alternatives. In fact, agents neglect the distinctions that there may be between two states "1" because they are only interested in experimenting with a state that is different from "0".

It follows that the number of true choice alternatives that are randomly changed when decision makers experiment a new policy choice should be an important contingency affecting the efficacy of representation search strategies. In what follows, I will refer to the number of true alternatives that are randomly changed as the degree of noise affecting feedback signals.

To illustrate, consider again an aeroplane designer whose simplified representation includes a policy category for "window appearance". This simplified category aggregates, for instance, five true choice alternatives given by the total surface area, the number, thickness and material of glass layers, and whether the outer shape has sharp corners. In this example, the representational accuracy of this category is assumed fixed at this level of aggregation (five). The designer wants to experiment a different window design. Feedback is characterised by low noise when a different window choice only changes one or few of the true choice alternatives. This may occur, for instance, when the designer accepts as a valid alternative for experimentation any design that is mostly equal to the original one and that differs in only one or a few features. However, by definition of simplified representation, the designer neglects or is unaware of at least some of the distinctions among the features that have been changed. This implies that while only a few features have changed with respect to the old design, some of them will be implemented randomly at each replication of the window because they are not under the direct control of the designer. For instance, the designer may select a new window appearance that only differs in terms of a larger surface area. As other designers are in charge of integrating the new larger window within the structure of the fuselage, they may need to adjust one of the windows' outer shape to make it compatible with some specific location of the fuselage's structure – e.g., one of the emergency exits. This adjustment may involve the inclusion of a sharp corner, which before the Comets fatal accidents would have been seen as an inconsequential modification.

In contrast, feedback is characterised by high noise when a different policy choice changes most or all the associated true choice alternatives. With reference to the previous example, this may occur when the designer accepts as a valid alternative for experimentation only window designs that look very different from the original along all features. As the new design is implemented, several design true choice alternative are neglect and out of the control of the designer, and implementation of the finer-grained details will vary randomly at each replication of the design.

Accordingly, I replicate the results of Figure 3 but for different levels of feedback noise. Specifically, the previous results were obtained by changing a random number of true choice alternative at each policy search step. In this set of results, agents change only one random true choice alternative at each policy search step when feedback noise is low, and all but one true alternative when noise is high.

Figures 4 and 5 show the results for low and high feedback noise, respectively. Evidently, feedback noise has a large impact on the performance of narrow and broad representation search strategies. The first result to note is that the performance of any search strategy, including the performance of the two benchmark agents, is significantly higher when feedback noise is small – approaching to 1 (the global peak) in Figure 4 for low complexity and the Fully Accurate agent. This result occurs because agents always avoid distant search regardless of degree of simplification of their representations.

Secondly, for low feedback noise, broad search strategies always perform better than intermediate and narrow strategies for all levels of complexity. This result can be explained in terms of the amount of residual randomisation subsequent to each representation search step. There are two randomised elements under low noise conditions, the selection of the policy category to search and of the true choice alternative. Narrow search breadth produces, on the one hand, one accurate dimension along which policy categories aggregate a progressively smaller number of true choice alternatives. On the other, it retains nine highly simplified dimensions, each representing one policy category that aggregates 16 true alternatives. It follows that when agents perform a policy search step and choose a policy category at random, they are more likely to choose one of the refined categories that aggregate fewer true alternatives. In turn, for low feedback noise, only one of those few true alternatives can be chosen for experimentation. Thus, all else being equal, agents try fewer true alternatives for narrow breadth than for broad breadth and performance is inferior because they experimented fewer policy configurations.

The results of Figure 5 show that the opposite is true when feedback noise is high. That is, narrow search strategies perform better than broad strategies for almost all levels of complexity, except for very low-complexity environments. Following the above line of argument, this result obtains because agents always perform distant search – i.e., they change a large number of true choice alternatives at once – when noise is high, except when a policy category is fully accurate and represents only one true choice alternative. Akin to the results of Figure 3, Panels C and D, it is beneficial to focus representation search efforts along one or fewer dimensions when search is distant because choice alternatives tend to be interdependent with multiple other choices. This argument is also supported by the results of Figure 5, Panels A and B. When complexity is low and true choice alternatives are less interdependent, distant search is less detrimental, and it is more beneficial to use broader representation search strategies.

Searching for optimal representations

The final set of results aim to determine the optimal degree of representational complexity and whether the likelihood of achieving optimal representations depends on different representation search strategies. Note that I previously defined the optimal representation for a given decision environment as the set of representations that maximise the performance attainable via policy search. The main result of this section is that simplified representations can outperform more accurate ones, including the fully accurate representation that perfectly matches the true performance landscape, and maximize the performance attainable by policy search. However, this is true only for broader search strategies ($b \geq 6$), while the optimal representation for intermediate ($b = 5$) and narrow ($b \leq 4$) search strategies is the fully accurate representation.

These results are obtained as follows. I run simulations for narrow ($b = 1$) and broad ($b = 10$) representation search strategies according to the baseline search process defined in the previous section, with two modifications. First, representation search does not stop when the policy category of the narrow search strategy achieves the highest level of accuracy, that is, after 15 representation search steps. Instead, representation search continues until agents' representations achieve the highest level of accuracy across all the 10 policy categories, that is, after 150 representation search steps. This allows me to determine whether optimal representation exists at all levels of representational accuracy. For the narrow search strategy ($b = 1$), full representational accuracy can be achieved by selecting a new policy category when the current one has been refined to the highest level of accuracy. That is, the narrow representation search strategy proceeds by refining to the highest level of accuracy one policy category at the time until full representational accuracy is achieved. In contrast, a broad representation search strategy proceeds by picking and refining one of the 10 policy categories at random at each step until full representational accuracy is achieved.

Second, after each representation search step and for both breadth strategies, agents perform repeated policy search iterations until a performance peak is reached and steady state is achieved. The steady state performance value is then compared to the steady state performance value that is obtained according to the baseline search process defined in the previous section; that is, by alternating representation and policy search steps until full representational accuracy and a performance peak are achieved. The difference between the two steady state performance values indicates whether an intermediate degree of complexity exists at which stopping representation search produces higher steady state performance than the steady state performance achievable with a fully accurate representation.

The results are show in Figure 6 for five values of environment complexity (K). The above process produced a steady state performance value for each level of representational accuracy of agents' representations, which is indicated in the x-axis, and for narrow ($b = 1$) and broad ($b = 10$) representation search strategies. The y-axis shows the difference between these values and the steady state performance value that is obtained by achieving the fully accurate representation with the respective representation search breadth strategy.

The results confirm that broad search strategies (red lines) can generate less accurate representations that lead to higher steady state performance values than more accurate ones, including the fully accurate representation. In contrast, for narrow search strategies (green lines), less accurate representations lead to lower performance peaks than more accurate ones for all values of complexity and representational accuracy. This holds for all intermediate to low values of representation search breadth, i.e., for $b \leq 5$.

Note that, however, the performance difference between fully accurate and less accurate representations is low ($< 1\%$) and constant even for intermediate and narrow search strategies. Interestingly, the level of representational accuracy at which the difference is lower than 1% decreases as the complexity of the environment K increases. For instance, for intermediate values of complexity $K = 80$ and for $b = 1$, the representational accuracy at which representations are nearly optimal is nearly 100 - that is, 66% of the highest level of accuracy.

In contrast, for broader search strategies ($b \geq 6$), both the levels of representational accuracy at which the optimal representation is achieved and at which performance equals the performance of the fully accurate representation increase as the complexity K increases. Further, in line with the baseline results of Figure 3, the difference between the steady state performance values of the optimal and fully accurate representations decreases as the complexity K increases.

DISCUSSION

This study examines the performance implications of different representation search breadth strategies when decision-makers' initial representations of a decision task are over-simplified and search efforts result in greater representational accuracy. The breadth of representation search shapes policy search depending on two contingencies, namely the complexity of the decision task and the noise of feedback signals. In the baseline case of random feedback noise, broad representation search is beneficial at low levels of complexity, narrow search is beneficial at intermediate and moderately high levels of complexity, and search breadth does not affect performance for very high levels of complexity. However, broad representation search is beneficial at all levels of complexity for low feedback noise, while narrow representation search is beneficial at all levels of complexity for high feedback noise.

This study contributes to the search and learning literatures by offering a reinterpretation of local peaks of performance in complex decision environments. While complex environments can indeed be described as rugged landscapes characterised by multiple local peaks of performance (Levinthal, 1997), there is often a mismatch between what decision makers believe is a local peak and actual local peaks of the landscape. That is, decision makers may often stop their process of adaptation via local search at locations of the landscape that do not correspond to local optima. Akin to "sticking points" that emerge due to organisations' internal decision structures (Rivkin, Siggelkow, 2002), "apparent peaks" of performance result in suboptimal adaptation. Their emergence, however, is due to the almost inevitable fact that decision makers' adaptation is guided by their simplified representations of the performance landscape (Simon, 1990; Levinthal, 2011). Specifically, apparent peaks are a consequence of the fact that decision makers do not have control over finer-grained features of the environment that their simplified representations neglect but that they nonetheless affect by means of local policy changes. In this sense, this study extends literature in this stream that explores the performance implications of different representations of dimensions of performance (Csaszar and Levinthal, 2016) and of interdependencies between policy choices (Martignoni, Menon, and Siggelkow, 2016) by examining the implications of the degree of accuracy with which policy choices are represented.

A practical implication of this finding is that experimenting via innovative and distant configurations of policy choices is not necessarily the only way to improve over or "escape" local performance optima.

What managers perceive as a configuration of choices where all the possibilities for further improvements have been exhausted may actually be an apparent local peak of performance. Thus, rather than by distant search, opportunities for improvement may be discovered more cheaply in the neighbourhood of current configurations by refining the representation of the decision problem in use. It is important to note, however, that this study has assumed perfect or ideal representation search efforts. That is, representation search always resulted in two finer-grained categories of policy choices that perfectly capture the distinctions of the underlying environmental features. Indeed, this assumption is most likely unrealistic in organisational settings, where managers or expert advisors may conjecture inconsequential or arbitrary refinements of policy choices. In fact, as I argued in the previous chapter, the capability to draw meaningful distinctions and devise more sophisticated but useful representations might be one of the central features of superior decision competence and domain expertise.

A second contribution of this study is to suggest that noisy feedback signals are not necessarily a consequence of non-deterministic characteristics of the environment but also due to how decision makers implement policy changes and interpret their outcomes by means of simplified representations. While apparent local peaks are a consequence of the fact that multiple finer-grained features of the environment are affected at once when decision makers experiment and implement new policy changes, noise is a consequence of the randomness with which finer-grained features are selected in an uncontrolled way, and increases with their number. The emergence of noise as a consequence of simplified representations is analogous to scientific laboratory settings where the experimenter does not have perfect control over the experimental environment. A critical requirement of laboratory settings is that a same configuration of initial conditions must always produce the same system responses, and this can occur only if the experimenter has perfect control over all the features of the environment that may affect these outcomes. Noisy response signals are thus observed when the experimenter, akin to the simple-minded manager, changes initial conditions and makes unintended changes to the system feature over which she has no control[18]

[18] Besides measurement errors and time-varying unobservable causes (Pearl, 2009).

In line with the motivation for this study, these results also contribute to the literature on knowledge specialisation (Teodoridis, Bikard, and Vakili, 2019). As discussed, the distinction between specialists and generalists is rooted in a strategic trade-off. Knowledge professionals can either invest their limited time across several domains and achieve a more superficial understanding of each; or acquire a deeper and more refined understanding of one or few domains. Research has found inconsistent evidence of the performance advantages of one type of knowledge specialisation over the other. For instance, specialised scientists (Leahey, 2007) and inventors (Conti, Gambardella, and Mariani, 2013) can be more successful as they are able to identify very specific and highly consequential gaps in their domains. Nonetheless, generalist scientists (Schilling and Green, 2011) and inventors (Reagans and Zuckerman, 2001) span broader and more distant knowledge domains can produce more creative recombinations. The study by Teodorisis, Bikard and Vakili (2019) suggests one way of reconciling these findings by identifying the pace of change in a knowledge domain as a key contingency. In particular, they show that generalists perform better when the pace of change is slow while specialists have advantages in fast-changing environments.

The findings of this chapter suggest that the complexity of the environment is an equally important contingency affecting the performance of knowledge specialisation. Generalists benefit from less deep but broader understandings of the environment at low levels of complexity, while specialists benefit from narrow and deep understanding of one or few domains at intermediate to moderately high levels of complexity. These results can be interpreted from the perspective of knowledge recombination and depth. When complexity is low, the performance contributions of distant knowledge domains do not depend on the configurations of other domains. In this conditions, even moderate levels of refinement, attained via representation search, and of experimentation, attained via policy search, are likely to result in a broader range of domains that each has a higher-than-average contribution to performance. In contrast, representations where one domain is deeply refined and highly performing, and the others provide average performance contributions, are expected to have a lower overall performance in low complexity environments. On the contrary, as complexity increases, environments demand increasingly higher levels of specialisation across all their domains that generalists are less likely to achieve. The relative advantage of being a specialist in at least one area of expertise increases in these environments.

More broadly, this study shows that the breadth of representation search is an important contingency affecting the performance of dual representation and policy search strategies. A counterintuitive result is that, in line with recent findings in this research stream, less accurate representations can outperform more accurate ones – i.e., that the optimal degree of representational complexity does not necessarily match the true complexity of the environment (Csaszar and Levinthal, 2016). However, this study shows show that less accurate representations can outperform more accurate ones only for broad rather than narrow representation search strategies. Thus, these findings contributes to the literature on the role of mental representation in decision-making, which presents conflicting views of the effects of representational complexity and is divided between proponents of fast-and-frugal heuristics (Gigerenzer and Goldstein, 1996; Sull and Eisenhardt, 2015), of highly complex and accurate representations (Kiesler and Sproull, 1982; Weick, Sutcliffe, and Obstfeld, 1999), and of representations that match the complexity of the environment (Ashby, 1956). The implication for this research is that performance does not only depend on representational complexity per se, but also on how a given level of complexity is attained by means of different representation search strategies and of how complexity is distributed across the dimensions of a decision problem.

Finally, these findings have implications for the discovery of superior strategies. Behavioural perspectives of strategy argue that cognitively distant strategies offer superior performance because they are less likely to be discovered and thus competed away (Gavetti, 2011). Although traditional search models do not directly account for the effects of competition, they indirectly capture this view by showing that global performance peaks are difficult to reach due to the presence of local peaks. Specifically, the trajectories of agents adapting via local policy search whose initial locations on the landscape are sufficiently distant from the global peak are most likely attracted and most likely end at local peaks of the landscape. This study offers a perspective to further characterise what *distant* superior strategies might mean. That is, distance does not only depend on whether agents' initial configurations are closer to local or global peaks, but also on the degree of representational complexity that guides adaptive choices and of the representation search strategy used to attain it.

FIGURE 1

Example of five representation search iterations for $N = 24$ and $N_p = 3$, and for narrow search breadth $b = 1$

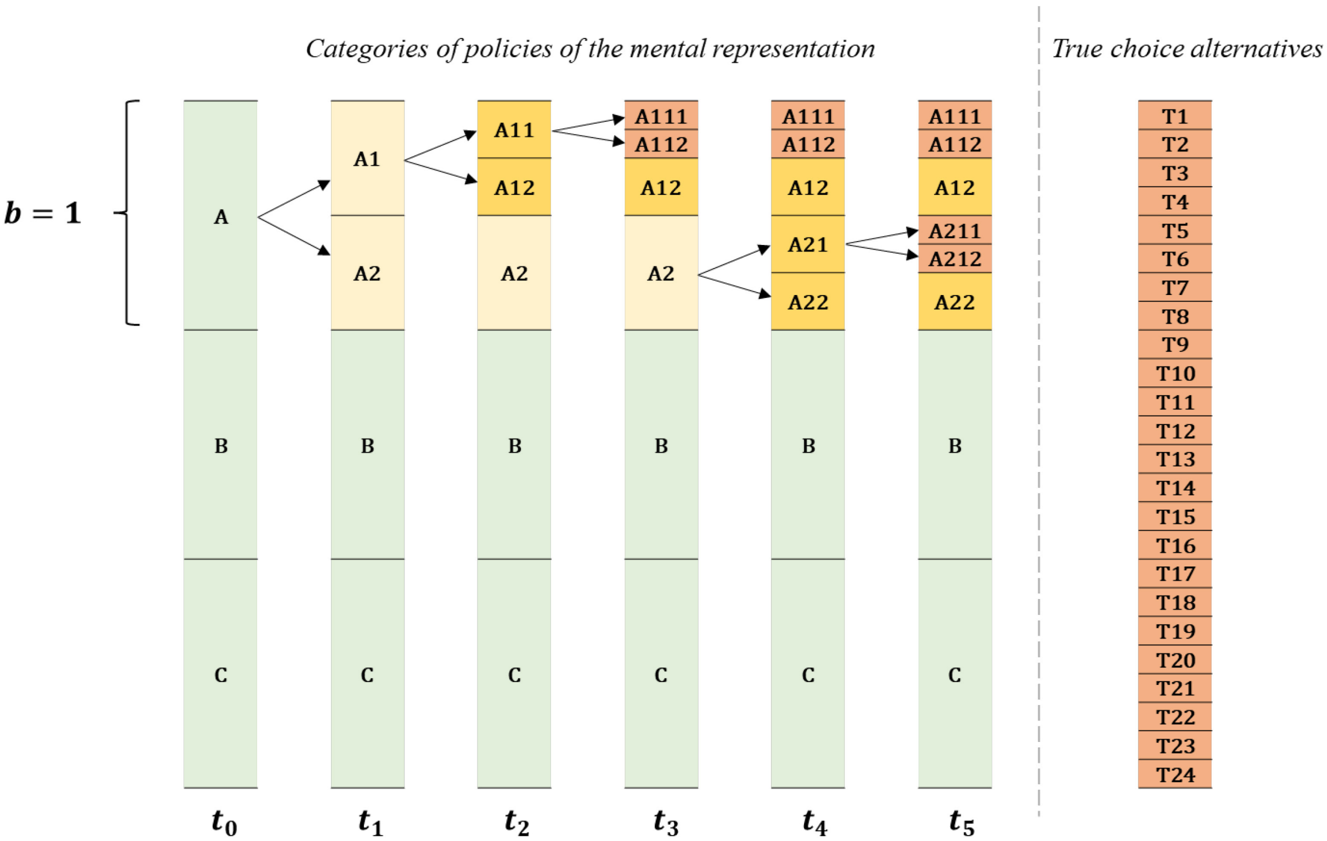


FIGURE 2

Example of five representation search iterations for $N = 24$ and $N_p = 3$, and for broad search breadth $b = 3$

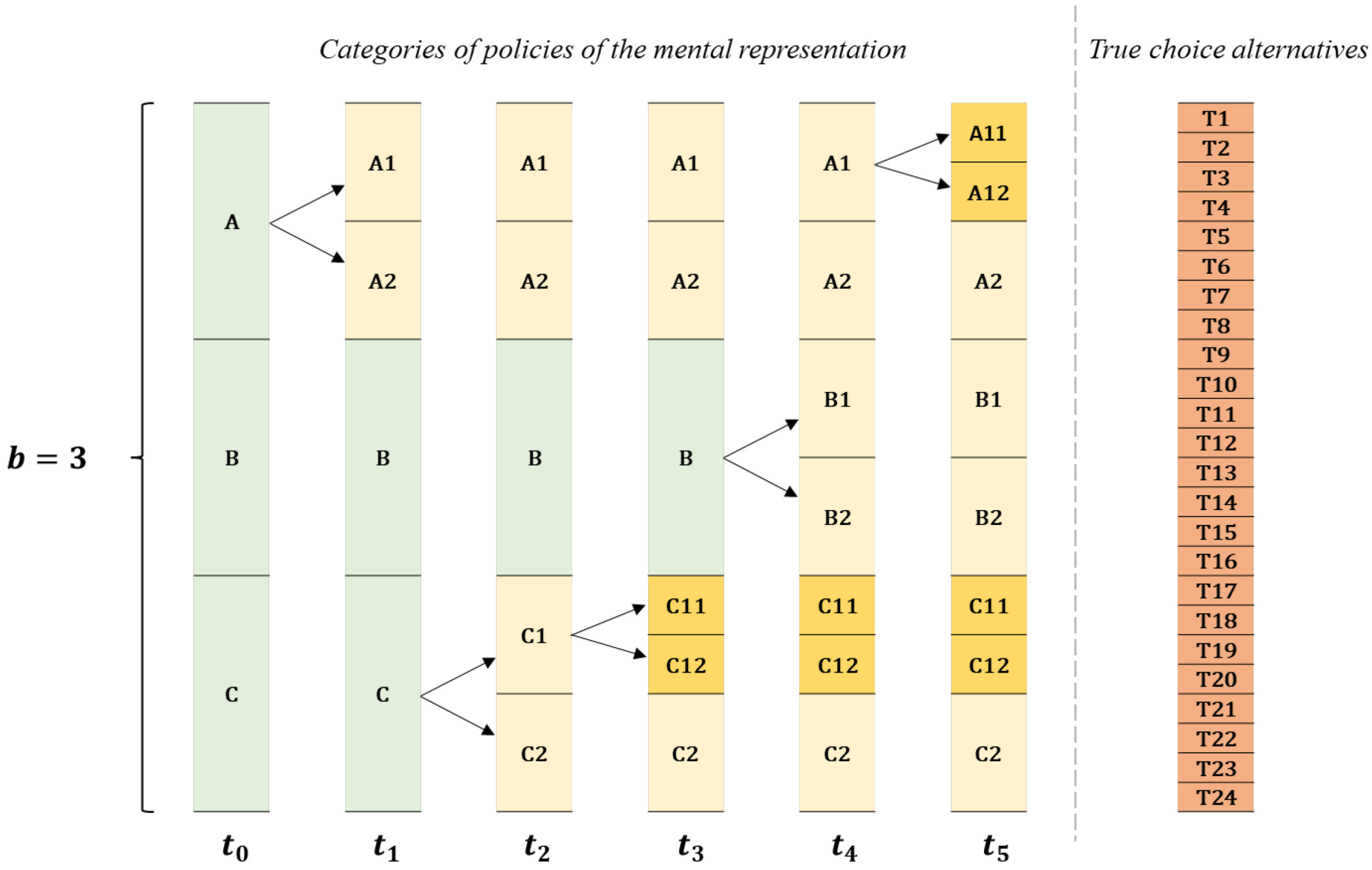


FIGURE 3

Baseline results. Performance over time of representation and policy search for different values of complexity, $K \in [15, 40, 80, 110]$ and for narrow ($b = 1$), intermediate ($b = 5$) and broad ($b = 10$) representation search strategies.

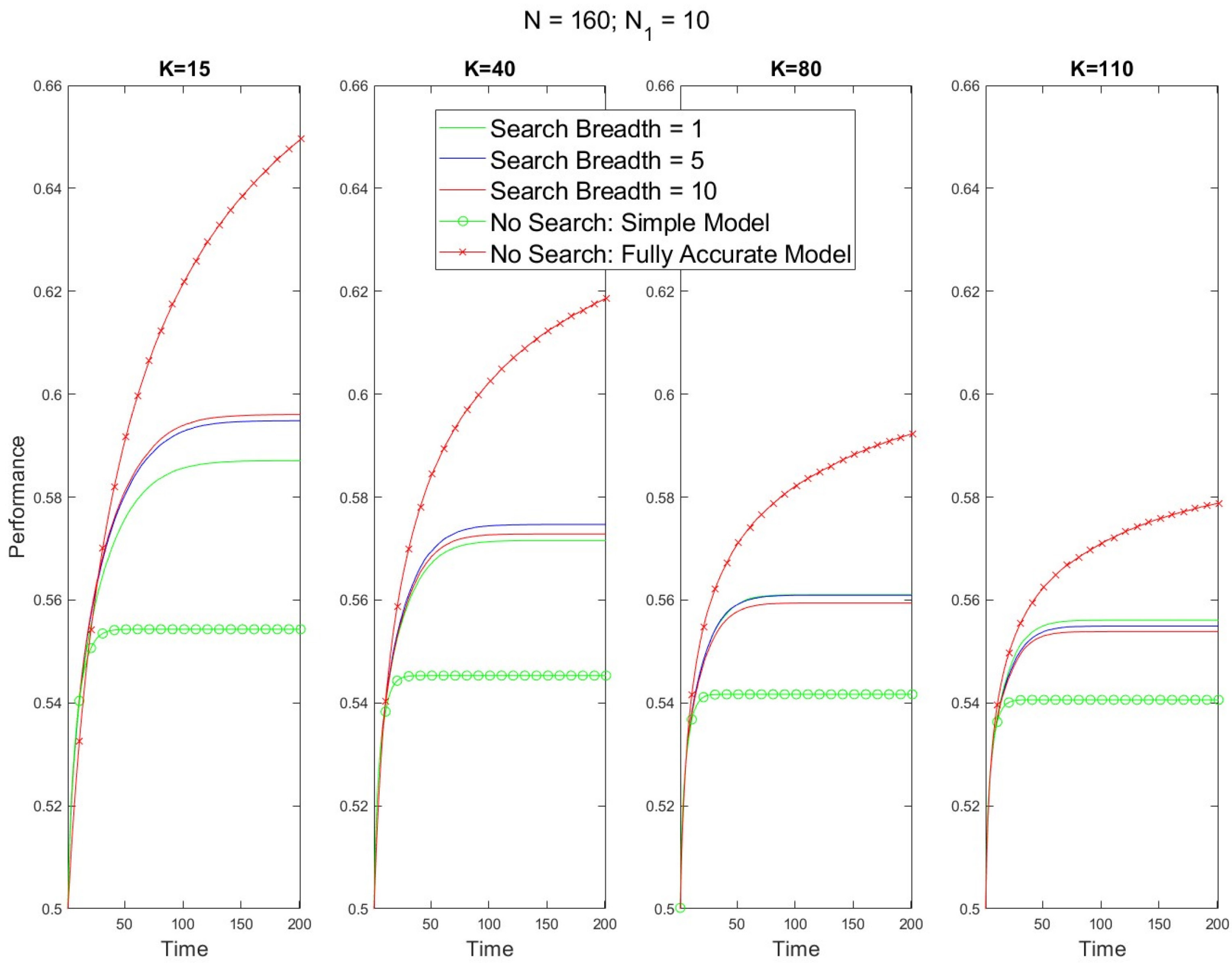


FIGURE 4

Effect of low feedback noise. Performance over time of representation and policy search for different values of complexity, $K \in [15, 40, 80, 110]$ and for narrow ($b = 1$), intermediate ($b = 5$) and broad ($b = 10$) representation search strategies.

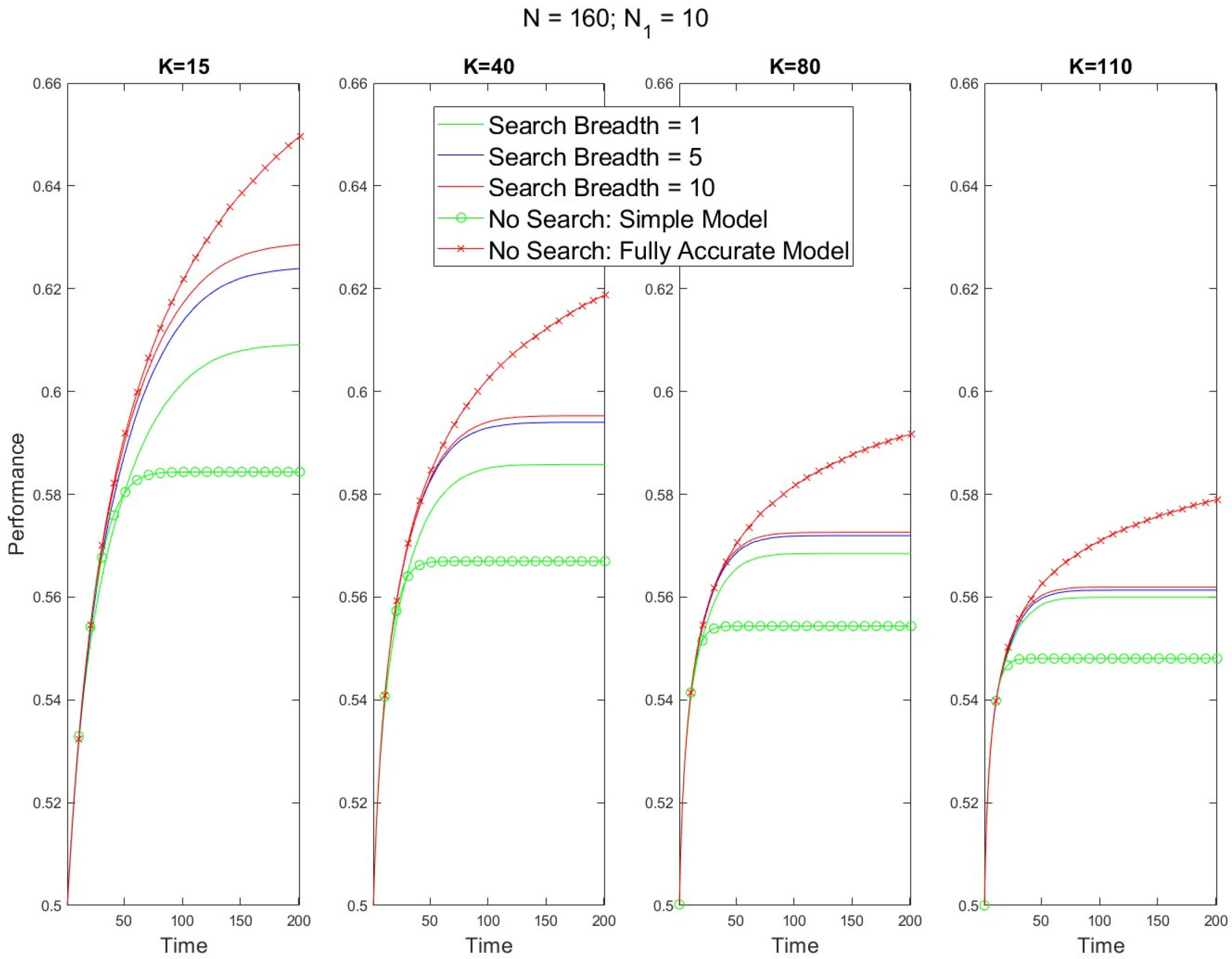


FIGURE 5

Effect of high feedback noise. Performance over time of representation and policy search for different values of complexity, $K \in [15, 40, 80, 110]$ and for narrow ($b = 1$), intermediate ($b = 5$) and broad ($b = 10$) representation search strategies.

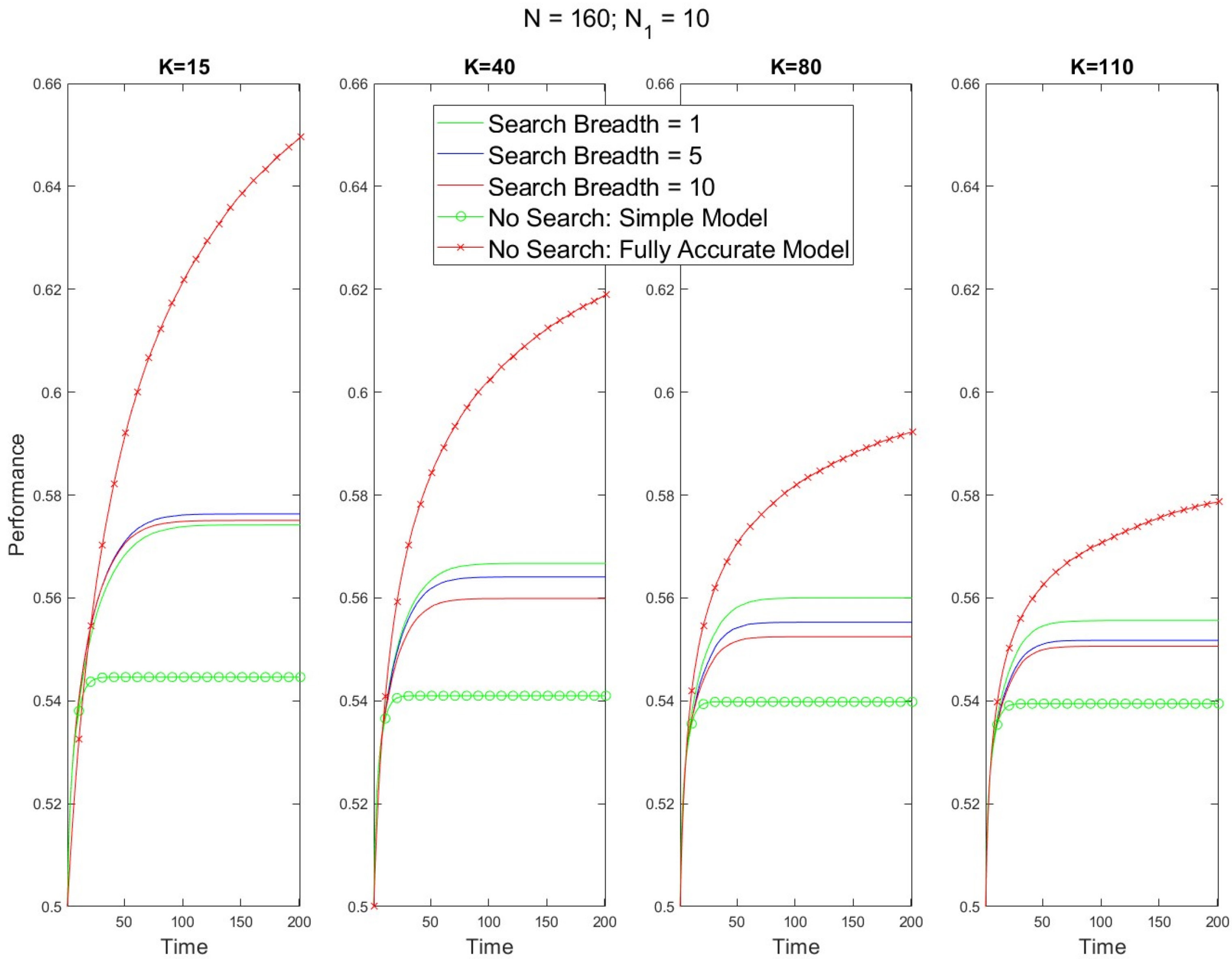
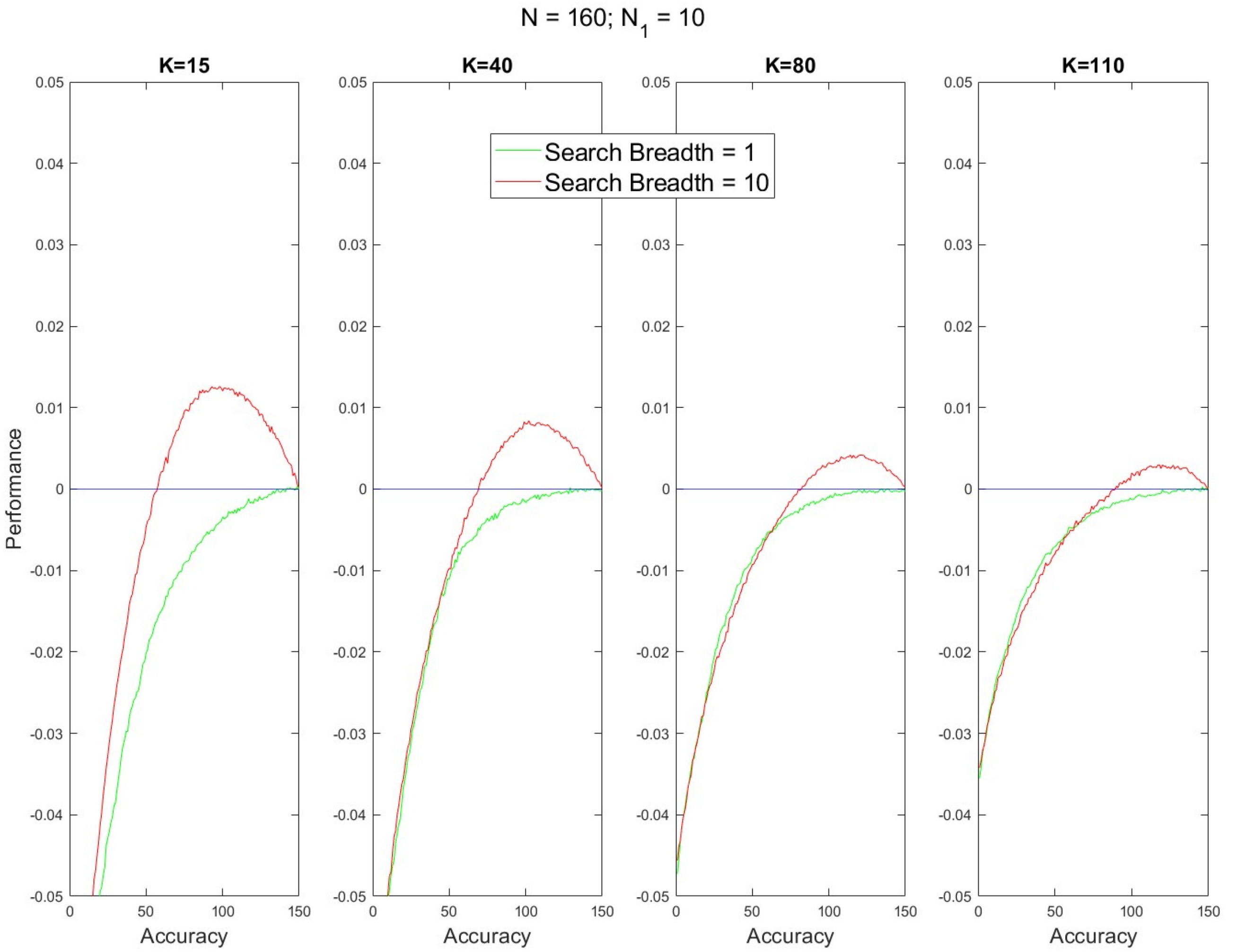


FIGURE 6

Difference between the steady state performance of representations with varying degrees of representational accuracy and the steady state performance of the fully accurate representation. Results are shows for different values of complexity, $\boldsymbol{K} \in [\mathbf{15}, \mathbf{40}, \mathbf{80}, \mathbf{110}]$ and for narrow ($\boldsymbol{b} = \mathbf{1}$) and broad ($\boldsymbol{b} = \mathbf{10}$) representation search strategies.



CHAPTER 3

GROUP EVALUATION ACCURACY: THE ROLE OF EXPERTISE DEPTH AND BREADTH IN THE SELECTION OF TECHNOLOGIES

ABSTRACT

This study explores the contingencies relating evaluators' domain expertise to the accuracy of group evaluations of technology, focusing on the diversity of expertise among evaluators. Research generally assumes that higher diversity of expertise improves the quality of decisions by providing access to a broader range of knowledge and information. While group expertise diversity is generally characterised in terms of differences between evaluators' distributions of expertise across technology areas – i.e. the group's *breadth diversity* – I introduce the group's *depth diversity* – i.e. evaluators' differences between their level of expertise in the technology being evaluated – as an equally important dimension of diversity. Results from empirical tests of a large dataset of group evaluations of patents made by expert evaluators at a Fortune 500 ITC firm support my arguments and offer two primary findings. I find that groups composed of evaluators who all had either high or low expertise in the focal technology – i.e. low depth diversity groups - were less likely to evaluate the value of patents accurately than groups comprising evaluators with both high and low levels of expertise – i.e. high depth diversity groups. Secondly, this positive effect of depth diversity on group accuracy was contingent on breadth diversity. That is, the positive effect of depth diversity on accuracy was lower for groups composed of evaluators who also differed in their expertise across technology areas – i.e. for groups characterized by both high depth and high breadth diversity. However, high breadth diversity improved accuracy for low depth diversity groups. These findings support my main argument that the role of evaluators' expertise in group evaluations of technology also depends on depth diversity as a distinct dimension of group expertise diversity. One implication for the literature on group evaluations and decisions is that, contrary to the general assumption, higher diversity of expertise does not improve decision quality when diversity is high on both depth and breadth dimensions.

INTRODUCTION

Technology evaluation is a key process for firms as it informs what early-stage technologies should be selected for commercialization. Yet, the value of embryonic technologies depends on multiple and interdependent dimensions of quality (Wang and Hsieh 2015; Higham, de Rassenfosse, and Jaffe, 2021), each of which can only be assessed under conditions of uncertainty (Freeman and Soete, 1997). Firms use various practices to reduce the uncertainty inherent in technology evaluations, including portfolio approaches (Khanna, Guler, and Nerkar 2016; Adner and Levinthal, 2004), stage gates (Cooper, 1990), and group decisions processes (Csaszar and Eggers, 2013; Brodbeck et al., 2007).

Group evaluations and decisions reflect one of the core functions of organizations as they facilitate internal exchanges of knowledge and information among members. Leveraging the expertise of multiple individuals is costly but can improve the overall quality of decisions (Brodbeck et al., 2007; Csaszar and Eggers, 2013). Firms commonly rely on multiple evaluators via committees, panels, or expert boards, among other forms (Li, Rosen and Suen, 2001; Criscuolo et al., 2017; Criscuolo et al., 2021; Hackman, 1990; Sundstrom, De Meuse and Futrell, 1990), or they encourage evaluators to seek advice from formal or informal networks of experts (Borgatti and Cross, 2003; Nebus, 2006).

However, while research generally assumes that aggregating the knowledge of multiple evaluators improves the accuracy of evaluations, the boundary conditions of this assumption remain understudied. This study examines one of the central tenets of group evaluations, that is, that diversity of expertise provides a wealth of knowledge and perspectives that collectively improve evaluation accuracy. This argument, generally associated with the "wisdom of crowds" logic (Surowiecki, 2004), rests on the assumption that knowledge gaps and individual biases cancel out in the average when multiple evaluations are aggregated.

While this assumption likely holds for large panels or when a large number of evaluations is crowdsourced, I explore its validity in other common organizational settings that only involve a limited number of evaluators who contribute to group evaluations through exchanges of information and group discussions. The implication for these settings is that diversity of expertise among evaluators may not necessarily aggregate in fruitful ways. Aggregation may be detrimental to evaluation accuracy because, contrary to situations in which the individual biases of a large pool of evaluators are averaged and cancelled out (Surowiecki, 2004), the aggregated accuracy of smaller groups is more sensitive to individual biases and evaluators can be more easily influenced by others' perspectives in group discussions. In line with this argument, a large number of empirical studies have found negative, positive, or even null effects of expertise diversity on various group decision outcomes (Miller et al., 2022)

I address this issue and derive the conditions under which aggregation of expertise can be detrimental to evaluation accuracy by noting that diversity among evaluators should be characterized across two distinct dimensions of expertise. Research has traditionally examined expertise diversity in terms of differences between individuals' distributions of expertise across knowledge domains, that is, the group's breadth diversity (e.g. Taylor and Greve, 2006; Criscuolo et al., 2017). This focus on breadth, commonly measured in terms of differences across categories of experience or functional backgrounds, has been motivated by the common assumption that group evaluations benefit from a diversity of perspectives and opinions (Miller et al., 2022)

I argue, however, that group breadth diversity is insufficient to characterize the effects of diversity of expertise and that the group's depth diversity plays an equally important role. In the context of technology evaluations, a group's depth diversity represents the extent to which evaluators differ in their level of expertise in the focal technology[19] (Mannucci and Yong,

[19] A group's depth diversity is thus a function of evaluators' individual expertise depth, defined as the individual level of expertise in the technology being evaluated (Mannucci and Yong 2018). High depth diversity implies

2018). Breadth diversity induces variation in perspectives derived from a diverse range of knowledge domains. Yet, there is no assurance that these perspectives will be biased in opposite directions, especially when the group size is limited (Brodbeck et al., 2007). Further, research suggests that individuals are more likely to discount or misinterpret perspectives and information from unfamiliar knowledge domains (Szulanski, Cappetta, and Jensen, 2004; Rader, Larrick, and Soll, 2017), undermining the potential contributions to accuracy of others' expertise.

Differences in the levels of expertise in the focal technology can instead benefit the accuracy of group evaluations because they induce perspectives that are systematically biased in opposite directions. I will argue that evaluators' level of expertise is not only associated with advanced and deeper knowledge but also with the likelihood of overestimating or underestimating the value of that technology (Dane, 2010; Boudreau et al., 2016). It follows that evaluators with different expertise in the focal technology contribute to group evaluations with perspectives and opinions that are systematically biased in opposite directions.

My main argument is thus twofold. First, the role of expertise diversity in group evaluations importantly depends on the group's depth diversity. Specifically, high depth diversity benefits accuracy because evaluators are biased in opposite directions and their biases tend to cancel out in group discussions. By the same token, low depth diversity reduces accuracy because it promotes the reinforcement of individual biases. Second, the effect of depth diversity on evaluation accuracy is contingent on breadth diversity[20]. When breadth diversity is high, evaluators have different interpretations of the focal technology and may fail to understand and integrate others' perspectives (Rader et al., 2017, Gavetti and Warglien, 2015). It follows that

that a group comprises similar numbers of evaluators with high and low levels of expertise in the focal technology. On the contrary, low depth diversity implies that almost all evaluators have either high or low levels of expertise in the focal technology.

[20] I note that breadth and depth diversity are independent dimensions of group expertise diversity. Evaluators may have acquired expertise across very similar sets of knowledge domains and yet have either large or no expertise in the technology being evaluated.

breadth diversity attenuates the positive effect of depth diversity on accuracy because it reduces the mutual attenuation of biases in group discussions. On the other hand, high breadth diversity can also reduce the mutual reinforcement of biases and thus be beneficial to accuracy when depth diversity is low.

I find support for my conjectures in the context of patent evaluations and termination decisions at a Fortune 500 high-tech firm. The firm's evaluators, or patent engineers, periodically evaluated all patents in the firm's portfolio and decided whether to renew or terminate them, depending on the patents' forecasted economic prospects. The firm's standard evaluation procedure recommended patent engineers rely on group evaluations by requesting the contributions of other evaluators, including the firm's technology experts and patent inventors, when they deemed it necessary. Although patent engineers were normally individually responsible for evaluations, nearly 25% of all patent evaluations were made by groups comprising between two and 10 expert evaluators. This dataset comprises nearly 20,000 written evaluations made by 140 patent engineers for 7,000 patent families between 1996 and 2016, including numerical ratings assigned to families at each evaluation and email exchanges between patent engineers and other evaluators. I measure evaluation accuracy as the difference between evaluations' ex-ante forecasts and patents' commercialization outcomes, which I can observe ex-post for each patent family and allow us to measure both overestimation (Type I) and underestimation (Type II) errors.

This study contributes to research on group evaluations of technology and group decision processes within firms. A key implication of these findings is that aggregating more and more diverse knowledge from multiple evaluators does not necessarily improve the accuracy of evaluations as generally assumed. I introduce an overlooked dimension of a group's expertise diversity – i.e. depth diversity - and show that group accuracy is reduced when diversity is high on both depth and breadth dimensions. A second implication is that, counterintuitively, groups of evaluators who all specialise in the focal technology would benefit not from the contributions of other experts of the focal technology but from evaluators with less or no expertise.

THEORETICAL BACKGROUND

Forecasts of the value of early-stage technologies are made under conditions of uncertainty. In this study, technological uncertainty refers to the unpredictability of the future development of a technology and of its economic prospects and builds on the notion of "environmental uncertainty" proposed by Packard, Clark and Klein (2017, p. 3). In line with this notion, evaluations are made under uncertainty because the outcomes of the future development of a technology belong to an open set of possibilities that are impractical or impossible to determine ex-ante. It follows that evaluators cannot assign probabilities to future events that may help them determine the economic prospects of a technology[21] (Knight, 1921). The open set of possible outcomes depends on the intrinsic characteristics of the technology (Fleming and Sorenson, 2004) and on external environmental factors including the co-evolution of closely related technologies (e.g., Kapoor and Furr, 2015), the creation of new business models (e.g., Zott and Amit, 2007), fluctuations in demand and user adoption (e.g., Adner and Levinthal 2001, Rogers, 1995), the competitive environment (e.g., Toh and Kim, 2013), and institutional and regulatory change (e.g., Van de Ven and Garud, 1993).

It is important to distinguish between subjective and true technological uncertainty (Packard, Clark and Klein, 2017). The characteristics of a technology and of the environment determine the true or objective level of technological uncertainty. In turn, evaluators have a subjective perception of the true level of uncertainty that depends on their individual expertise in the technology and experience with the decision environment (Konlechner and Ambrosini, 2019). We should expect that more accurate subjective perceptions of the true level of uncertainty

[21] The literature refers to decision-making under risk when outcomes belong to a closed set and can be assigned knowable probability distributions.

allow evaluators to make more accurate evaluations, at least on average[22]. Greater domain expertise, i.e. greater experience and knowledge of a domain is generally associated with a greater understanding of how future outcomes may unfold in that domain and should thus afford evaluators with more accurate perceptions of the true level of uncertainty (Chi, Feltovich and Glaser, 1981; Dane, 2010).

This distinction emphasizes the fact that even evaluators with expertise in a technology may often make inaccurate evaluations. Evaluators may fail to learn from past experiences or build overconfidence in their own subjective judgements over time, and thus develop a substantial gap between perceived and true uncertainty (Kruger and Dunning, 1999; Zollo, 2009). Further, true uncertainty is generally high for early-stage technologies and difficult to characterize. This fact is illustrated for instance by the patenting context, where several evaluation criteria and dimensions of quality are commonly used to evaluate patented inventions that can only provide approximate estimates of their overall value (Wang and Hsieh, 2015; Higham, de Rassenfosse, and Jaffe, 2021). Thus, technology evaluations remain challenging even for experienced individuals and firms.

Firms broadly use two approaches to manage uncertainty and improve the accuracy of evaluations. On the one hand, firms manage the risks associated with unpredictable outcomes by adopting diversification and contingency strategies (Milliken 1987, Adner and Levinthal, 2004; Teece, Peteraf, and Leih 2016; Khanna, Guler, and Nerkar 2016). On the other, they attempt to characterize uncertainty by reducing the set of future outcomes they need to take into consideration, for instance by ruling out outcomes that seem implausible (Roberts and Lattin, 1991). Firm's knowledge and individuals' expertise are crucial for this purpose, as they determine beliefs about the future development of a technology and the likelihood of future environmental states (Packard, Clark and Klein, 2017). Evaluators who do not possess

[22] Indeed, this assumption holds for any given level of true uncertainty only in relative terms and all else being equal, i.e. when comparing evaluators with accurate vis-à-vis inaccurate perceptions of true uncertainty.

knowledge and expertise in a technology area are mostly unaware of possible outcomes and would consider known outcomes as equally likely. On the contrary, evaluators with expertise possess an understanding of causal relationships between the characteristics of a technology and the external environment (Mosakowski, 1997). This knowledge allows them to form subjective judgements as to which outcomes are most likely rather than implausible and enhances the ability to forecast desirable outcomes (Einhorn and Hogarth, 1986; Csaszar and Otler, 2022). Put it differently, evaluators' expertise reduces the degree of perceived uncertainty from a level of total ignorance to a level that reflects more closely the true uncertainty of the technology environment.

Accordingly, firms use group evaluation processes to leverage the expertise of multiple individuals and improve evaluation accuracy (Csaszar and Eggers, 2013). As individuals develop different types of expertise because of specialization and division of labour, one of the core functions of organisations is to coordinate and facilitate internal access to advanced knowledge in different areas (Cyert and March, 1963; Thompson, 1967; Ren and Argote, 2011). Group evaluation processes reflect this function and leverage a broader pool of expertise to improve the collective understanding of how a technology might evolve in the future (Tindal, Kameda, and Hinsz, 2003). For instance, professional service firms commonly rely on panels of evaluators with diverse expertise to evaluate projects and make resource allocation decisions (Criscuolo et al., 2017).

It is thus important to examine the contingencies that determine the effectiveness of group evaluations in aggregating evaluators' unique expertise and improving evaluation accuracy. Evaluating in groups rather than individually demands additional time and resources that vary with the number of evaluators involved and with the evaluation process design (Thompson, 1967; Sah and Stiglitz, 1986). An understanding of how superior accuracy can be achieved in groups would thus allow to optimize the benefits-to-costs ratio of group evaluations.

To this purpose, I define evaluation accuracy as the difference between the estimated or forecasted and the true or realised value of a technology. Evaluation accuracy is low when evaluators overestimate (make Type I errors) or underestimate (make Type II errors) the future value of a technology. Overestimation leads firms to select technologies that will eventually generate less value than they forecasted, potentially resulting in disappointing returns or losses. For instance, home TV manufacturers that invested heavily in developing 3D screens likely overestimated the demand for this technology. On the contrary, underestimation occurs when firms fail to select technologies that will eventually generate higher value than they forecasted, potentially resulting in a severe competitive disadvantage. For instance, TV manufacturers who missed out on LCD panel technologies and invested instead in OLED and plasma panels found themselves at a competitive disadvantage with other market players (Eggers, 2012)

Thus, I examine the assumption that aggregating the expertise of multiple individuals would diminish the frequency of Type I and Type II errors. I take on an information aggregation view of evaluation and decision-making to analyse how individual evaluations may contribute to group evaluation accuracy (Cyert and March, 1963, pp. 19–22; Thompson, 1967). Firms use different schemes or decision-making structures for aggregating individual evaluations, including averaging, voting, or delegation to one or multiple specialised experts (Owen and Grofman, 1986). The central idea that underpins these structures is that aggregating a larger and more diverse pool of expertise provides a broader range of information and perspectives on the advantages and disadvantages of a technology that reduces the impact of individual knowledge gaps and biases (Hollenbeck et al., 1995).

The wisdom of crowds phenomenon illustrates how group evaluations are expected to improve evaluation accuracy according to this view. Individual evaluations may present some degree of positive or negative bias, that is, they may either over- or underestimate the value of a technology. The magnitude of individual bias depends on multiple factors, including evaluators' lack of knowledge and experience, overconfidence, or personal motives and preferences (Kahneman and Tversky, 1979). Individual biases tend to vary in opposite directions when multiple evaluations are aggregated, and individuals differ in their knowledge and experience. When the number of individuals is sufficiently large and the pool of expertise sufficiently diverse, biases are expected to cancel out in the average and the resulting aggregated estimate or forecast tends to the true or realized value or outcome (Suroweicki, 2004).

However, this rationale is less likely to hold when group evaluations are made within firms, for two reasons. First, cost limitations suggest that the number of individuals involved in group evaluations is relatively limited, especially when evaluations are numerous and frequent. The smaller the number of evaluators, the larger the influence of their individual biases on the overall evaluation accuracy. Thus, individual biases are less likely to cancel out in the aggregate as the number of evaluations gets smaller.

Second, group decisions in organisations are often reached as the result of collective discussions rather than by direct averaging or via other algorithms for aggregating independent evaluations (Cyert and March, 1963). Evaluators are not necessarily required to provide an overall quantitative rating or estimate – e.g. a value from a scale or interval – or binary vote - e.g. "accept/reject" – that can be directly used in aggregation schemes such as averaging or voting. Instead, they commonly engage in exchanges of opinions and group discussions by providing more qualitative assessments that other evaluators need to interpret and reconcile with their own perspectives (Boccaccio and Dalal, 2006). It is in this sense that group decisions in organisations function as a vehicle for exchanging, combining and integrating information from individuals with unique expertise with the purpose of accessing a richer knowledge base and building on each other's perspectives (Brodbeck et al., 2007).

These features of group decisions in firms have two implications. First, evaluators can be influenced by others' biases as a result of interactions and exchanges of opinions. The wisdom of crowds phenomenon is more likely to occur not only when groups are larger, but also when individual evaluations are formulated with limited interactions among evaluators (Suroweicki, 2004). Similar to blind peer review in academia, the latter condition ensures that individual evaluations are formulated independently and increases the likelihood that individual biases will cancel out in the average[23] (Lee et al., 2013). Experts within firms may instead have a significant influence on others' perspectives as they engage in group discussions (Stasser, Kerr and Davis, 1989). Research shows that people think their estimates are not influenced by others' estimates when in fact they are (Sherif, 1935; Nolan, Schultz, Cialdini, Goldstein, and Griskevicius, 2008) and that they underestimate the influence they have on others (Bohns and Flynn, 2013). Hence, interactions among evaluators may skew the distribution of individual estimates and prevent that individual biases cancel out in the aggregate.

Second, the benefits of aggregating the unique expertise of multiple evaluators depend on how they interpret and integrate others' perspectives. This point is illustrated by the distinction between surfacing and processing information in group decisions (Martins and Sohn, 2022). While diversity of expertise likely results in a broader range of information being surfaced and exchanged in group discussions, there is no assurance that information will also be processed in a constructive way so that individual biases can be exposed and eliminated. Brodbeck and colleagues (2007) discuss group- and individual-level processes that affect how information is processed in groups and that limit the extent to which individuals' perspective can be integrated. For instance, information shared by a larger number of individuals is judged more credible and important than conflicting information shared by a minority of individuals regardless of the objective truth of the minority perspective (Stasser and Titus, 1985).

[23] A further analogy can be made in statistics for the mean of a random sample, which is an unbiased estimate of the true value of a population mean as long as elements of the sample, i.e. individual evaluations, are drawn randomly and independently.

Finally, I note that these effects of interactions among evaluators can have a large impact on evaluation accuracy regardless of the aggregation scheme used for their individual evaluations. Different aggregation schemes can be more or less effective in reducing the impact of individual biases depending on group size and on how biases are distributed among evaluators (Tindale et al., 2003, Laughlin, 2011). However, if a subset of evaluators were highly biased against or in favour of a technology and could influence most of the other evaluators' perspectives, the aggregated accuracy would reflect the systematic shift of individual biases regardless of whether averaging, voting or other schemes were used to aggregate opinions (Einhorn et al., 1977).

In the next section, I will focus on technology evaluation contexts where the number of evaluators is limited and interactions are important for evaluation accuracy. These conditions reflect common organisational settings and allow me to examine how different distributions of evaluators' expertise contribute to the effective aggregation of knowledge.

HYPOTHESES: THE ROLE OF EXPERTISE DIVERSITY

My main argument is as follows. Diversity of expertise plays a critical role in improving the accuracy of group evaluations. However, I argue that the degree of expertise diversity among evaluators must be characterized not only in terms of *breadth diversity* as traditionally assumed, i.e. in terms of differences between evaluators' distributions of expertise across technology areas, but also in terms of the group's *depth diversity*, i.e. of differences between evaluators' levels of expertise in the focal technology. The reason is that a group's breadth and depth diversity have two distinct effects on group evaluations of limited size and where interactions are important. A group's depth diversity affects the extent to which individual evaluations are systematically biased in opposite directions, while a group's breadth diversity affects the extent to which other experts' perspectives are integrated rather than discounted. I argue below that the degree to which group evaluations improve evaluation accuracy is contingent on the interplay between both dimensions of expertise diversity.

It is worth noting that depth and breadth diversity are indeed two distinct and independent dimensions of expertise diversity. With reference to Figure 1, evaluators' expertise can be described in terms of their total volume of experience and of how their experience is distributed across technologies. The total volume of experience distinguishes novices (Figure 1, Panel a) from expert evaluators (Figure 1, Panel b). Evaluators are expected to learn from experience, and it is reasonable to assume that experts make relatively more accurate evaluations than novices, on average (Levitt and March, 1988; Csaszar and Ostler, 2020).

In fact, this study will focus exclusively on expert evaluators and on the implications for group evaluations of different distributions of expertise among them. In the following, I will refer to expert evaluators or simply evaluators interchangeably. Similarly, I will refer to evaluators with expertise as a shorthand for evaluators with expertise in the technology being evaluated.

<<Insert Figure 1 about here>>

With reference to Figure 1 Panel (b), low breadth diversity implies that evaluators have similar distributions of expertise across the same technologies and is observed in two situations. Consider two pairs of evaluators C, D and E, F (bottom left panel). The breadth diversity of both two-member groups is low because both C, D and E, F have similar distributions of experience across the five technologies in the example. These two groups are also characterised by low depth diversity because evaluators C, D and evaluators E, F have similar experience with the focal technology (technology 3 in the example). In particular, evaluators C, D have large stocks of experience and thus high levels of expertise in the focal technology, while evaluators E, F have limited expertise in the focal technology.

Other combinations of distributions of expertise are equally possible and independent from each other. The top right panel of Figure 1 Panel (b) illustrates the situation where groups are characterised by both high depth and high breadth diversity, while the off-diagonal panels are examples of the two other possible combinations of diversity. High depth diversity implies that evaluators have different levels of expertise in the focal technology (top panels), while high breadth diversity implies that evaluators' stocks of experience have different distributions across the same technologies or are distributed across different technologies (right panels).

These distinctions are meaningful because depth and breadth diversity affect group evaluations in different ways. The discussion of the previous section shows that while breadth diversity ensures access to a broader range of information and perspectives, it cannot guarantee on its own that the aggregated evaluation will be more accurate. For instance, consider a group comprising evaluators with expertise in multiple and different technology areas, that is, for whom breadth diversity is high. They will likely contribute to group discussions with a broader range of perspectives and arguments against or in favour of the technology being evaluated. Arguments will likely differ in terms of their assumptions, domain specific information and logical rationales, which depend on evaluators' unique expertise and are influenced by their subjective assessments. However, if the group size is limited, there are no reasons to expect that different arguments will also be against or in favour of the focal technology in a balanced proportion that ensures that the group is not collectively biased in either direction. Further, even if we could assume that different arguments were biased in opposite directions in a balanced proportion, there would be no reasons to expect that evaluators will interpret and be influenced by them in a way that reduces their biases when they have expertise across different areas (Brodbeck et al., 2007, Martins and Sohn, 2022).

I will argue that, contrary to breadth diversity, group depth diversity is systematically associated with the degree to which individual evaluations are biased in opposite directions. As I mentioned, this fact warrants the emphasis on the distinction between the two dimensions of expertise that I propose below.

I develop my main argument as follows. First, I argue that evaluators with expertise in a technology are positively biased in favour of that technology (**H1a**) and negatively biased against technologies they know less (**H1b**). Second, group depth diversity induces a mutual attenuation of individual biases and is thus associated with higher evaluation accuracy (**H2**). Finally, I show that this positive effect of depth diversity on evaluation accuracy is contingent on breadth diversity (**H3**).

The association between expertise depth and positive bias in evaluations can be explained as follows. First, evaluators are subject to the familiarity bias, a cognitive bias in which individuals tend to favour known choice alternatives and of which they have previous experience (Kahneman and Tversky 1979). This bias is well documented in the finance literature, as investors show a large preference for domestic as opposed to foreign investments (French and Poterba 1991; Huberman 2001) and for stocks they know well, regardless of performance (Biais, Hilton, Mazurier and Pouget 2005) This influence of familiarity on individual preferences and choices can be seen as a form of risk aversion. In the context of technology evaluations, technologies with which evaluators have limited experience are perceived as more uncertain and, all else being equal, are evaluated less positively (Fox and Tversky 1995, Boudreau et. al. 2016). On the contrary, evaluators will show lower risk aversion and will prefer technologies with which they have greater experience and that they know better.

Second, evaluators who specialise in a technology may have personal interests in promoting that technology, especially in the context of portfolio evaluations where different technologies compete for investments. Evaluators may want to ensure the survival and relevance of a technology among competing others because they want to protect the relevance and value of their own expertise within the firm. In other words, evaluators may show an "agency bias" as they believe that their reputation and competence within the firm depend on the relevance of the technology for which they are internally recognized as experts. By the same token, evaluators may be negatively biased against technologies in which they do not specialise as they compete for the same resources.

Third, despite possessing deeper and higher quality knowledge, evaluators with expertise in a technology are nonetheless influenced by familiarity and agency biases because of overconfidence in their own subjective judgements. It would be reasonable to object to the above arguments by arguing that evaluators with expertise possess a more comprehensive understanding of the reasons for both the success and failure of the focal technology and that they should be more accurate despite the influence of other biases. I argue instead that the above biases would still influence evaluators with expertise because high levels of true uncertainty make subjective judgements important and subjective judgements are driven by overconfidence.

Specifically, an implication of high levels of true uncertainty is that it is difficult to quantify the relative importance of all the possible reasons that would determine either the future success or failure of a technology. Although evaluators with expertise in that technology may be aware of those reasons, they still need to rely on their subjective judgement to decide which set of reasons prevails on the other and formulate an estimate of the most likely future outcome (Knight, 1921, Packard et al 2017). Thus, familiarity and agency biases influence evaluators with expertise by influencing their subjective assignment of weights to reasons for success rather than for failure. Overconfidence, in turn, plays a role in amplifying the influence of these biases. Research shows that individuals tend to build overconfidence with experience, especially in tasks and areas that they perceive as familiar (Kahneman and Tversky 1979; Glaser, Langer, and Weber 2007). Thus, although unsure of objective and quantifiable rationales for assigning higher weights to reasons for success rather than failure, evaluators with expertise will have confidence in their biased subjective assessments and formulate overly positive judgements.

In sum, while expertise furnishes a more comprehensive understanding of a technology, it also boosts experts' overconfidence in their subjective assessments, which are influenced by familiarity and agency biases. For instance, Hall, Ariss and Todorov (2007) offer experimental evidence that individuals with more information and knowledge on a topic have greater overconfidence in their own judgments. When asked to provide estimates pertaining to that topic, participants with more information were less accurate than less informed individuals as they used information to confirm their pre-existing subjective beliefs and preferences (Hall, Ariss and Todorov 2007).

In contrast, a lack confidence in their subjective assessments and limited knowledge of the focal technology have the opposite effect on expert evaluators with less expertise in the technology, who tend to be negatively biased. As I argued above, lack of relevant knowledge and higher perceived uncertainty induce more negative subjective judgements due to risk aversion (Fox and Tversky 1995, Boudreau et al 2016). Evaluators' lack of confidence in their own assessments likely increases their perceptions of uncertainty and thus exacerbates their aversion to risk.

Finally, it is worth noting that positively (negatively) biased evaluations are associated with higher (lower) likelihood of overestimation and lower (higher) likelihood of underestimation. For instance, evaluators may be required to assign ratings to a portfolio of R&D proposals that their firm will use to prioritize investments. Positively biased evaluators will systematically assign higher ratings controlling for proposals' quality and other proposal characteristics. Hence, ratings that are systematically higher all else being equal will necessarily result in more frequent overestimation and less frequent underestimation of the future returns or outcomes of the proposals. The opposite holds for negatively biased evaluators. For these reasons, I posit:

H1a: Evaluators with greater depth of expertise in a technology are more likely to overestimate its value.

H1b: Evaluators with greater depth of expertise in a technology are less likely to underestimate its value.

In turn, evaluators can influence other evaluators' subjective judgements. As discussed, the mutual influence that evaluators can have on each other is one of the central features of group evaluations within firms. The literature on group decisions-making identifies two motives that explain the mutual influence among participants, namely normative versus informational influence (Deutsch and Gerard 1955). Under normative influence, individuals tend to conform with, satisfy or support others in order to gain social approval (MacGeorge, Feng, and Guntzviller 2016) or to avoid potential confrontations and rejections (Wood 2000). In contrast, under informational influence, individuals' judgements change as they learn new information and perspectives from others and update their beliefs accordingly (Ecken and Pibernik 2016). Both motives emphasize the importance of exchanges of information and interactions among evaluators in changing their own subjective judgements as a result of being exposed to the judgements of others (Rader, Larrick and Soll 2017). In fact, as mentioned, individuals would think their estimates are not influenced by the estimates of others when in fact they are (Nolan et. al. 2008) and they would underestimate the influence they have on others (Bohns and Flynn 2013).

The above discussion sets the stage for the first part of my argument that differences among evaluators in their level of expertise in the focal technology have a systematic effect on the accuracy of group evaluations. When depth diversity is low and expert evaluators have similar degrees of expertise with the focal technology, individual biases tend to mutually reinforce and to increase their influence on group accuracy. On the contrary, when depth diversity is high, individual biases tend to cancel out and their influence on group accuracy is diminished.

Before I examine the relationship between depth diversity and evaluation accuracy, it is worth comparing depth diversity with the effects of breadth diversity discussed in the previous section. I previously argued that there are no reasons to expect that evaluators with different distributions of expertise will make subjective judgements of a technology that are biased in opposite directions. While breadth diversity is associated with the range of different information and arguments that evaluators surface in group discussions, the way in which they use their unique information and knowledge to support their subjective positive or negative judgements depends primarily on their level of expertise. This distinction is important because, contrary to what it is generally assumed, it is not breadth diversity but rather depth diversity that has a systematic effect on the mutual influence between individual biases and thus on the accuracy of group evaluations.

The effect that depth diversity has on the mutual reinforcement and attenuation of biases depends on the confidence that evaluators have on their subjective judgements and on the influence exerted by other evaluators' judgements. To illustrate, I first examine the case in which depth diversity is low as evaluators all have high expertise in the focal technology. As previously argued, evaluators with expertise possess more and more detailed information and knowledge of the focal technology, including knowledge of different causal paths or rationales for its future success or failure (Fiske and Taylor 1984, Dane 2010). Despite their deeper knowledge, even evaluators with expertise must rely on their subjective judgement to assess the importance of each causal path and formulate their own assessments of the technology because true technological uncertainty is generally high (Knight, 1921). This fact has two implications. Although, as previously discussed, evaluators with expertise develop overconfidence and a positive bias in favour of the focal technology, they are still aware of rationales and potential arguments against its future success. This awareness partially limits the degree to which overconfidence and positive biases would make them support the technology if they evaluated it individually.

Secondly, evaluators must rely on their subjective judgement also to assess other evaluators' information and arguments. That is, familiarity and agency biases will also influence how evaluators filter and interpret others' evaluations. This effect thus compounds to the fact that the arguments that surface in group discussions are already filtered and positively framed themselves under the influence of familiarity and agency biases. The result is a further increase in evaluators' confidence in their biased assessments. These biases are thus mutually reinforced by other evaluators with expertise and the aggregated evaluation will be even more positively biased.

The opposite reinforcement occurs when depth diversity is still low, but evaluators do not have expertise in the focal technology. In this case, evaluators possess less detailed information and more superficial knowledge of causal paths and rationales for the success or failure of the technology, and they are relatively less confident about their subjective judgements. Risk avoidance and a tendency to favour other competing technologies in which evaluators have competence result in a negative bias towards the focal technology. Nonetheless, self-awareness of possessing more superficial knowledge in this area partially limits the degree to which evaluators with no expertise would discount the technology if they evaluated individually. Similar to the previous case, this negative bias will both influence which arguments are surfaced in group discussions and how others' arguments are interpreted by evaluators. Additionally, risk aversion due to unknown outcomes is further reinforced when evaluators realise that none of them has more detailed information and knowledge of the technology. Hence, akin to the previous case, individual biases are mutually reinforced when depth diversity is low, although in this case the aggregated evaluation will be more negatively biased.

The effect of high depth diversity on group accuracy follows from the previous two cases. When depth diversity is high, evaluators with both high and low expertise contribute to group discussions and support arguments that are respectively positively and negatively framed. While research suggests that individuals tend to partially discount distant perspectives held by others and will not be fully influenced by them (Eagly and Chaiken 1984), evaluators' individual biases are nonetheless attenuated as a consequence of the exposure to others' opposite opinions (Brodbeck et al 2007). Evaluators with expertise tend to be overconfident and positively biased but are aware that counterarguments exist. Although evaluators without expertise put forward counterarguments that might be less detailed and compelling, they still have an influence on the former, who are induced to reassess and adjust their judgements. Indeed, an equivalent adjustment occur for evaluators without expertise, who are also less confident of their knowledge of the focal technology and are more likely to trust the expertise of others.

The result is that individual biases tend to be attenuated or cancel out when the group's depth diversity is high, improving group evaluation accuracy. While the overall effect on the aggregate evaluation depends on the balance between the number of evaluators and their level of expertise, the influence of individual biases tends to diminish as depth diversity increases. Importantly, high depth diversity also ensures that at least some evaluators possess specialised knowledge of the focal technology, and that more detailed and compelling arguments are surfaced and considered in group discussions. This is important because the fact that individual biases are attenuated does not necessarily imply that the aggregated evaluation accuracy has improved. Despite being collectively less biased, evaluators may overlook important causal paths to the future success or failure of a technology, or be aware of only superficial ones that suggest incorrect conclusions. The attenuation of individual biases implies that these causal paths are assessed more objectively and used correctly in arguments that are surfaced in group discussions. This latter point is important because one may assume that an intermediate level of expertise exists at which evaluators are neither positively nor negatively biased. A group where all evaluators possess this optimal level of expertise would thus have low depth diversity and not be influenced by individual biases. However, higher depth diversity would still be more beneficial to accuracy because it ensures that important casual paths are considered in group discussions and that they are used in arguments more objectively. Therefore, I posit:

H2: There is a positive relationship between depth diversity among evaluators and the accuracy of group evaluations.

The second part of our argument is that the foregoing relationship is contingent on group breadth diversity. The key effect of breadth diversity when groups are of limited size and interactions are important is that differences between evaluators' distributions of expertise across technologies affect the degree to which they are influenced by the exposure to others' perspectives. All else being equal, larger differences between evaluators' distributions of expertise make it more difficult for them to interpret and integrate others' arguments and diminishes the influence that they have on others' perspectives.

Before I analyse this mechanism in more detail, it is worth emphasizing its implications for group accuracy, which are twofold. High breadth diversity reduced the benefits of high depth diversity because it reduces the mutual attenuation of individual biases. High breadth diversity is thus detrimental to accuracy when depth diversity is high. By the same token, however, breadth diversity also reduces the mutual reinforcement of individual biases when depth diversity is low. That is, high breadth diversity is beneficial to accuracy when depth diversity is low. As previously mentioned, this interplay between depth and breadth warrants a distinction between the two dimensions of expertise diversity and shows that higher breadth diversity is not necessarily beneficial when interactions are important.

The implications of breadth diversity for group accuracy are based on the effects of expertise depth and breadth at the individual level, which determine how evaluators understand and interpret arguments and opinions. As discussed, expertise depth is related to the amount and quality of information evaluators have on a technology, and to the comprehensiveness and level of detail of their knowledge of causal paths to its future success or failure. In turn, expertise breadth is related to the range of distinct knowledge domains or technologies in which evaluators have expertise. The important implication of individuals' expertise breadth is that it facilitates connections and analogies between domains of expertise. Having expertise in two distinct domains allows evaluators to interpret one area from the perspective of the other and to see interdependencies between them. Put it differently, evaluators would not be able to use their knowledge of a technology to understand another technology in which they don't have expertise.

In turn, the depth of expertise in each technology covered by one's breadth generates a unique distribution of expertise that determine whether evaluators are equipped to understand and integrate the arguments offered by others. Each distribution of expertise furnishes knowledge from a range of domains at different level of detail that individuals organise in their unique knowledge structures – i.e. in their representations of the knowledge contents pertaining to a domain and of the linkages between domains[24] (Dane, 2010). Knowledge structures determine how decisions are made, experience and past observations are organised, and information is processed and communicated (Bower and Hilgard 1981, Fiske and Taylor 1984). Importantly, this also includes a central role in informing the arguments and rationales with which evaluators contribute to group discussions and in interpretating others' arguments and rationales. In particular, interpreting arguments based on unfamiliar knowledge demand greater cognitive costs due to the tacitness of knowledge contents and to an insufficient understanding of the linkages between familiar and unfamiliar domains. Hence, evaluators with very different distributions of expertise and knowledge structures are less likely to assimilate and process the information and arguments that they provide to each other (Cohen and Levinthal 1990, Szulanski, Cappetta, and Jensen, 2004).

The key implication is that larger difference between evaluators' distributions of expertise reduce the extent to which they can influence or convince each other with their own arguments.

[24] Knowledge contents include information that has meaning specific to the knowledge domain, causal relationships and paths, and specific ways of processing information and causal relations, which are all acquired through experience or learned over time from others (Thagard 2010).

The additional cognitive costs needed to interpret unfamiliar arguments can have two consequences. On the one hand, research shows that individuals tend to discount information and advice that they cannot interpret and understand (Eagly and Chaiken 1984, Borgatti and Cross, 2003). On the other, they force evaluators to rely on subjective interpretations of information and arguments to fill in assumptions and causal connections they cannot fully understand. The implication for group accuracy is the same in both cases. When breadth diversity among evaluators is high, group discussions are less effective as a vehicle for the mutual adjustment of subjective judgements and for reaching a balanced collective interpretation of the technology (Gavetti and Warglien 2015).

For instance, a laptop manufacturer may have the opportunity to invest in multiple technologies, including display, moving parts such as hinges and locking mechanisms, manufacturing processes, software and CPU components. A group of evaluators characterized by high depth and breadth diversity and tasked with the evaluation of a display technology may include specialised experts with expertise in display, moving parts and manufacturing technologies and less specialised experts with expertise in software and CPU components. Besides being less positively biased towards investing in displays, the contribution of the different perspectives that the latter evaluators can bring to a group discussion is derived from their expertise in software and CPU technologies. Evaluators specialised in displays may be receptive to arguments based on knowledge of technologies that they *perceive* as adjacent and relevant to display – in this case, moving parts and manufacturing – while they may fail to see the connections with other areas.

Finally, we need to distinguish this effect of breadth diversity on group discussions in the two cases where depth diversity is respectively high and low. As discussed above, group accuracy benefits from high depth diversity because subjective judgements are influenced by both positive and negative biases that tend to cancel out through interactions and mutual adjustments. All else being equal, high breadth diversity creates greater barriers to mutual interactions and diminishes the mutual attenuation of biases. Thus, all else being equal, breadth diversity reduces the positive effect of high depth diversity on group accuracy. In contrast, high breadth diversity can be beneficial when the result of interactions among evaluators is to reinforce their individual biases – that is, when depth diversity is low.

Thus, I posit:

H3: Breadth diversity among evaluators negatively moderates the positive relationship between depth diversity and group evaluation accuracy.

EMPIRICAL CONTEXT: MANAGING THE PATENT PORTFOLIO AT ALPHA

To test my hypotheses, I require a context where expert evaluators in an organization: (a) evaluate different technologies to decide whether to continue or terminate ongoing investments; and (b) engage in evaluations with a limited number of other evaluators who can interact by sharing information and opinions. These conditions are satisfied in the context of the management of the patent portfolio of Alpha, a multinational, Fortune-500 ITC firm (pseudonym).

As many other large firms, Alpha regularly reviews all active portfolio patents to both identify opportunities for value creation and save on maintenance costs. The regular re-evaluation of patent portfolios is necessary because the potential value that a patent can generate changes over time, as new competing technologies are developed or new market opportunities arise (Guler 2007; Khanna, Guler, and Nerkar 2018). Patent rights can be renewed with the respective patent offices for up to 20 years, subject to the payment of recurring maintenance fees. Maintenance costs can be reduced by terminating patents, frequently by reducing the size of a given patent family rather than by terminating the entire family at once. Serrano (2010) estimates that nearly 50% of all patents are terminated before their legal term by their owners.

In order to better understand the patent evaluation process at Alpha, I conducted interviews with employees and with Alpha's director of IP, and analysed evaluation guidelines and other internal documents. I conducted one-hour interviews with 16 evaluators based in 7 locations worldwide and collected 10 responses from evaluators to a 15-question survey.

At Alpha, patent evaluations were performed by patent engineers, often with the contribution of other patent engineers, technology experts and patent inventors. Alpha granted individual decision authority over evaluations to patent engineers but encouraged them to rely on group evaluations with other expert evaluators, including patent inventors, as they deemed it necessary. Patent engineers also had discretion on whom to involve in group evaluations, which they likely decided depending on their workload, experience with the focal technology, knowledge of who had expertise in the area and on their availability, among other factors.

For each patent evaluation, patent engineers were responsible for writing an evaluation statement, assigning a rating, and terminating one or more family members as deemed necessary. Other evaluators were also formally required to write their own evaluation statements and, optionally, to provide rating and termination recommendations. The statements were meant to describe a patent's limitations and highlight opportunities for value creation that the firm could potentially exploit and to provide useful information for future re-evaluations. The patent engineers with decision authority coordinated group discussions via email and were responsible for assigning ratings and making renewal or termination recommendations that aggregated the group's perspectives. Ratings were numerical ranging from 0 (low) to 5 (high) according to broad evaluation guidelines. Patent engineers could also add identifiers to the numerical ratings to identify patents that could potentially become part of a technology standard or be implemented in Alpha's products.

Written evaluations and email exchanges among evaluators were stored in a software system to which patent engineers had access. The system was also accessible to Alpha's commercialization units, such as standardization, infringement and litigation, or the product implementation units, that used the ratings and the evaluation statements to guide their commercialization efforts.

Alpha's patents were categorized according to internal technology classes and grouped under distinct patent boards. Each patent board was responsible for managing the filing, prosecution and maintenance of the patents for its respective technology classes. Patent engineers could be assigned to the evaluations of patents from any patent board depending on workloads and upcoming prosecution and renewal deadlines, although some of them could specialise in one or few areas. This could happen especially as they developed an internal reputation for having expertise in certain technologies and other patent engineers sought their expertise in group evaluations of related patents.

The interviews confirmed that the decision to consult other evaluators was primarily driven by the fact that evaluations were generally highly uncertain. Alpha did not provide specific criteria or rules for making evaluations, and patent engineers had significant discretion over ratings and termination decisions. Internal documents and evaluation guidelines mention broad dimensions of quality or evaluation factors against which patents' future economic prospects could be evaluated, such as 'legal protection' or 'business value'. However, Alpha did not provide specific training or instructions on how the specific characteristics of patents could be assessed and mapped into these broad criteria. Evaluators relied instead on their own knowledge and subjective judgement. They admitted that it was often difficult to identify factors that could clearly indicate whether patents had potential for commercialization. Patent engineers also confirmed that these patents were given intermediate ratings (2 and 3) and that were the "*most difficult to evaluate*". This substantial level of uncertainty is reflected by the longstanding challenges of assessing patent value, as discussed in the patent literature (Wang and Hsieh, 2015; Higham, de Rassenfosse, and Jaffe, 2021).

The evaluation process at Alpha is thus suitable for testing my hypotheses. Patent engineers accumulate experience with a range of different technologies and acquire higher level of expertise in some of them. As evaluations are highly uncertain, they may leverage the expertise of other expert evaluators in group evaluations. These evaluators include other patent engineers, technology experts and patent inventors, who have their own degree of specialisation with the focal technology and unique distributions of expertise across other technologies. They are required to provide written evaluations and qualitative recommendations, thus their contribution to the evaluations depends on how individual arguments and perspectives are interpreted and integrated.

DATA AND METHODS

Data

I constructed a dataset with information drawn from Alpha's patent portfolio management system, matched with archival data on patents obtained from external sources. I collected nearly 18,000 written evaluation statements and ratings pertaining to approximately 9,000 patent families that Alpha filed between 1988 and 2006. Most of the evaluations were made within the first ten years of patents' life (92%), and each family was evaluated nearly three times on average. The statements were produced by 146 Alpha patent engineers, who each on average evaluated 161 patent families and made 150 evaluations between 1990 and 2016. Alpha used 166 3-digit technology classes to identify their patents. Patent engineers evaluated an average of 54 technology classes, up to 113, and made an average of 24 evaluations per technology class, up to 600.

My main sample consists of the nearly 4,000 (25%) evaluations made in group evaluations with other patent engineers, technology experts or patent inventors. These other evaluators, on average, contributed to 53 group evaluations; evaluated 19 technology classes, up to 88; and made 33 evaluations per technology class, up to 450.

I used Alpha and external data sources to collect information about successful patent commercialization outcomes, which include licensing agreements, sales, patent standardizations, litigations, and implementations in products. Alpha's commercialization units recorded dates and details of commercialization events. I complemented and validated this information with data collected from two external databases. I collected data on transfers of ownership and litigation legal actions from Google Patents, and additional data on litigation legal actions and infringements from Clarivate's Darts-IP database. More than 3,700 (23%) patents of the full sample and more than 800 (21%) of my main sample were successfully commercialized during the observation period. Commercialization occurred with approximately equal frequency between five and fifteen years after filing, while the rate of commercialization sharply decline sixteen or more years after filing.

Variables and Measurements

Overestimation, Underestimation and Accurate Evaluations

The variables describes whether patent engineers correctly estimated the future value of patents and accounts for both overestimation (Type I errors) and underestimation (Type II errors). I used changes in engineers' patent family ratings as indicative of positive or negative evaluations. The guidelines recommended, but did not prescribe, reducing or "*trimming*" the number of active members for a patent family with lower ratings (ratings 0 and 1), and renewing or expanding active members for families with higher ratings (ratings 4 and 5). Additionally, as previously mentioned, ratings guided commercialization units' search for valuable patents in the portfolio. Hence, ratings were meaningful indicators of the value that patent engineers attributed to patents, and ratings' increases or decreases reflected patent engineers' positive or negative forecasts of patents' future prospects[25].

[25] The average rating assigned to a patent was 2.54 (SD 1.01). Intermediate ratings 2 and 3 were assigned respectively, 42% and 38% of the times and ratings 1 and 4 were assigned respectively, 6% and 8% of the times. Nearly 65% of all evaluations confirmed the most recently assigned rating, whereas ratings were upgraded or downgraded by one unit respectively, 16% and 11% of the times and by two units nearly 4% and 3% of the times. Most upgrades by one unit occurred for families rated 2 (58%) and 3 (24%), while downgrades by one unit occurred for families rated 3 (55%) and 2 (27%).

Considering changes in ratings, I defined overestimation and underestimation in the following way. Overestimation occurs when patent engineers increased or maintained the ratings of patents they believed had positive future prospects and that subsequently expired without being commercialized; while underestimation occurs when they decreased the ratings of patents they believed had negative future prospects and that were subsequently commercialized. See Appendix A for an example of how I computed overestimation and underestimation.

Consequently, evaluations are inaccurate when either overestimation or underestimation occurs, and accurate otherwise. Specifically, evaluations are accurate when patent engineers increased or maintained the ratings of patents they believed had positive future prospects and that were subsequently commercialized; and when they decreased the ratings of patents they believed had negative future prospects and that subsequently expired without being commercialized.

This measure of evaluation accuracy has at least three desirable characteristics. It is an objective measure of the accuracy of patent engineers' estimates (Zollo, 2009). Second, it accounts symmetrically for both the objectives of the patent evaluation process at Alpha. That is, it captures the accuracy of evaluations in terms of identifying both high value patents for commercialization and low value patents for saving on maintenance and management costs. Finally, this measure avoids issues of manipulability (Zollo, 2009), because commercialization decisions did not depend on patent engineers but on the business units and other parties.

As a caveat, I cannot measure the accuracy of evaluations in which it was decided to terminate all the active members of a patent family, because it is not possible to know whether commercialization could have occurred had the patent engineers not decided to abandon these families. However, this issue has limited consequences in this setting because full abandonment decisions were made for less than 6% of the families in the portfolio.

I operationalize *overestimation* and *underestimation,* the dependent variables of H1, and *accurate evaluation*, the depend on variable of H2 and H3, as follows:

$$overestimation = \begin{cases} 1 & \text{if } assigned\ rating - previous\ rating \geq 0 \text{ and } commercialization = 0 \\ 0 & \text{otherwise} \end{cases}$$

$$underestimation = \begin{cases} 1 & \text{if } assigned\ rating - previous\ rating < 0 \text{ and } commercialization = 1 \\ 0 & \text{otherwise} \end{cases}$$

$$accurate\ evaluation = \begin{cases} 1 & \text{if } overestimation = 0 \text{ and } underestimation = 0 \\ 0 & \text{otherwise} \end{cases}$$

The variable $commercialization$ is equal to 1 if patents were eventually commercialized and equal to 0 if patents expired without being commercialized, while *assigned rating* and *previous rating* were, respectively, the rating assigned by the patent engineer as a result of the focal evaluation and the rating assigned to the patent as a result of the most recent evaluation[26].

Experience, Expertise Depth, and Expertise Breadth

These variables describe the depth and breadth of evaluators' expertise in terms of the experience they accumulated with Alpha's technology classes. As discussed, patent engineers, Alpha's technology experts and patent inventors contributed to group evaluations. I counted the distribution of evaluations made across Alpha's technology classes prior to the focal evaluation, including the evaluations that patent engineers made individually. For patent inventors, I additionally counted the inventions they disclosed internally prior to the focal evaluation to better characterize their expertise across technology classes. As I discuss below, the positive bias associated with specialisation tends to be even more pronounced for inventors as they are additionally subject to an "ideator's bias", i.e. they prefer their own inventions over competing ones (Fuchs et al 2017).

[26] Each patent entered Alpha's portfolio with a preassigned rating.

I measure *expertise breadth* by counting the number of technology classes in which an evaluator accumulated experience prior to the focal evaluation (Taylor and Greve, 2006; Mannucci and Yong, 2018).

I measure *expertise depth* as the proportion of experience accumulated by an evaluator in the focal technology class to the total experience accumulated by the evaluator prior to the focal evaluation. Formally, I indicate with $e_{i,t}$ the experience accumulated by individual i in the technology area t prior to the focal evaluation f. The total prior experience accumulated by individual i across all T Alpha's technology classes is:

$$experience_i = \sum_{t=1}^{T} e_{i,t}$$

Thus, the depth of expertise of individual i with respect to technology area t at the time of evaluation is given by:

$$depth_{i,t} = \begin{cases} \frac{e_{i,t}}{\sum_{t=1}^{T} e_{i,t}} = \frac{e_{i,t}}{experience_i} & \text{if } experience_i > 0 \\ 0 \quad \text{otherwise} \end{cases}$$

Alternatively, prior research has measured depth as the experience accumulated in each domain, i.e. technology class, $e_{i,t,f}$ (Mannucci and Yong, 2018). However, this measure emphasises the level of expertise in the focal technology relative to other technologies and is better suited to capture the influence of familiarity and agency biases. For instance, consider two evaluators who both evaluated technology A 50 times but have made respectively a total of 65 and 200 evaluations across all technology classes. Even if they have the same experience with technology A, class A constitutes respectively around 75% and 25% of the stock of experience of the two evaluators as the second evaluator accumulated most of her experience in technologies other than A. Even if class A was the class that the second evaluator evaluated the most, she would less likely prefer A for a lack of familiarity with other technologies and have less incentives to penalise other technologies in favour of A. The former evaluator, in contrast, would be more likely influenced by the perception that her expertise within the firm depends on the relevance of technology A.

In robustness tests, I limited both measures to the experience accumulated over a time period of 6 or 3 years prior to the evaluation or counted only technology classes for which at least 5 or 10 evaluations were made. Further, I measure experience breath by excluding the focal technology and using the Herfindahl–Hirschman index as an alternative operationalisation (Bunderson and Sutcliffe 2002, Lee and Csaszar 2017). These alternative sets of measures provide results that are consistent with my chosen measures.

Depth and Breadth Diversity

A group's breadth diversity measures the differences between evaluators' distributions of expertise across technology classes. I derive a variable based on the cosine difference between evaluators' expertise distributions (Criscuolo et al; 2017). The expertise distribution of evaluator i is a vector $\boldsymbol{x}_i$ of 166 elements, where each element is the number of evaluations of the corresponding technology class prior to evaluation f. The *breadth diversity of evaluator i* is the average cosine distance between i's expertise distribution and the expertise distribution of all the other $n-1$ evaluators j contributing to a group evaluation:

$$breadth\ diversity_i = \frac{\sum_{j=1}^{n-1}\left(1 - < \boldsymbol{x}_i, \boldsymbol{x}_j >\right)}{n-1}$$

where $< \boldsymbol{x}_i, \boldsymbol{x}_j > = \frac{x_i \cdot x_j}{\|x_i\| \|x_j\|}$ is the cosine similarity between $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$.

The group's *breadth diversity* is thus given by the average individual breadth diversity:

$$breadth\ diversity = \frac{\sum_i^n breadth\ diversity_i}{n} = \frac{\sum_i^n \sum_{j=1,j>i}^n \left(1 - < \boldsymbol{x}_i, \boldsymbol{x}_j >\right)}{n(n-1)/2}$$

which, as the formula shows, is bounded between 0 and 1. For instance, when patent engineers involve only one expert evaluator, breadth diversity is simply the cosine difference of the expertise distributions of the two evaluators.

A group's depth diversity measures the differences between evaluators' expertise in the focal technology. With reference to the individual variable $depth_{i,t}$ defined above, the *depth diversity of evaluator i with respect to technology t* prior to the focal evaluation is the average absolute difference between i's level of expertise in technology t and the degrees of specialisation in technology t of all the other $n-1$ evaluators j contributing to the group evaluation:

$$depth\ diversity_{i,t} = \frac{\sum_{j=1}^{n-1} \left| depth_{i,t} - depth_{j,t} \right|}{n-1}$$

This variable is bounded between 0 and 1 and approaches 1 when evaluator *i* has either a very high (close to 1) or very low (close to 0) level of expertise in technology *t* while all the other $n-1$ evaluators have respectively very low (close to 1) or very high (close to 0) expertise in the technology.

I measure the group's *depth diversity with respect to technology t* by computing the average individual depth diversity multiplied by a corrective factor $1/\widetilde{N}$:

$$depth\ diversity_t = \frac{1}{\widetilde{N}} \frac{\sum_i^n depth\ diversity_i}{n} = \frac{1}{\widetilde{N}} \frac{\sum_i^n \sum_{j=1,j>i}^n \left| depth_{i,t} - depth_{j,t} \right|}{n(n-1)/2}$$

where $\widetilde{N}$ is given by:

$$\widetilde{N} = \begin{cases} \dfrac{\frac{n-1}{2}\left(\frac{n+1}{2}\right)}{n(n-1)/2} = \dfrac{1}{2}\dfrac{n+1}{n} & \text{if } n \text{ is odd} \\ \dfrac{\left(\frac{n}{2}\right)^2}{n(n-1)/2} = \dfrac{n}{2(n-1)} & \text{if } n \text{ is even} \end{cases}$$

and ensures that $depth\ diversity_t$ is bounded between 0 and 1. For instance, when patent engineers consult only one expert evaluator, $\widetilde{N} = 1$ and depth diversity is simply the absolute difference between the individual level of expertise of the two evaluators. As the number of evaluators increases, $\tilde{N}$ tends to $1/2$ and so would do the upper bound of $depth\ diversity_t$ without the corrective factor $1/\tilde{N}$.

In fact, my measure of depth diversity can be seen as a measure of group polarization (Bramson et al 2022). Suppose that the individual variables $depth\ diversity_{i,t}$ are distinct and sorted in ascending order such that $depth\ diversity_{i,t} < depth\ diversity_{j,t}$ for each pair of indexes $i < j$. The group $depth\ diversity_t$ can be rewritten as:

$$depth\ diversity_t = \begin{cases} \dfrac{\sum_{i=1}^{\frac{n+1}{2}} (n+1-2i)(depth_{n+1-i,t} - depth_{i,t})}{\frac{n-1}{2}\left(\frac{n+1}{2}\right)} & \text{if } n \text{ is odd} \\ \dfrac{\sum_{i=1}^{\frac{n}{2}} (n+1-2i)(depth_{n+1-i,t} - depth_{i,t})}{\left(\frac{n}{2}\right)^2} & \text{if } n \text{ is even} \end{cases}$$

Suppose now that $n = 6$ and that there are two subgroups of evaluators with high and low expertise in the technology *t*, respectively equal to 0, 0.1, 0.2 for evaluators 1, 2, 3 and equal to 0.8, 0.9, 1 for evaluators 4, 5, 6. In this example, group depth diversity is:

$$depth\ diversity_t = \frac{5(1) + 3(0.9) + 1(0.8) - 1(0.2) - 3(0.1) - 5(0)}{9}$$

$$= \frac{5 + 2.7 + 0.8 - 0.2 - 0.3}{9} = \frac{8}{9} \approx 0.89$$

The high value of depth diversity indicates that this group is clustered or "polarized" at the extremes of the distribution of specialisation in technology *t*. The literature on opinion polarization in social groups discusses various measures of polarization, including dispersion, i.e. the average absolute difference between the individual measure and the group average, and spread, i.e. the difference between the group maximum and the minimum (Bramson et al 2022). My measure is similar to dispersion as it captures the extent to which evaluators' individual depth is different form the group's average; and it is closely related to spread, as it gives a larger weight (the factor 5 in the example) to the highest and lowest values of evaluators' individual depth.

In robustness tests, I limited both measures to the experience accumulated over a time period of 6 or 3 years prior to the evaluation, counted only technology classes for which at least 5 or 10 evaluations were made, and excluded the focal technology from the measure of breadth. Further, in line with prior work, I measured group breadth diversity by aggregating the experience of all the group members and counting the number of technology classes with which the group accumulated experience (Taylor and Greve, 2006). These sets of measures provide results that are consistent with my chosen measures.

Control variables

I included control variables pertaining to both evaluation and patent family levels. First, I included variables to account for different features across evaluations. The *uncertainty* variable accounts for the perceived technological uncertainty related to the focal evaluation. In line with the decision under uncertainty literature, I operationalize this variable as a linearly decreasing function of the focal patent's age, which I scale so that it ranges from 1 (high) to 0 (low) (Mosakowski, 1997; Konlechner and Ambrosini, 2019). Patents' age is the difference between the family's earliest filing year and the year at the time of evaluation. Age is inversely related to the residual uncertainty perceived by patent engineers because information on future value creation opportunities is disclosed over time and the likelihood of positive outcomes decreases over time (Mosakowski, 1997). I included the dummy variable, *flagged patent*, which equals 1 if the focal family was identified as potentially relevant to technology standards or product implementation in Alpha prior to the focal evaluation. To control for attention and cognitive capacity, I computed the number of evaluations made by the focal patent engineer in the previous seven days (*workload*). I constructed a dummy variable, *transferred case*, which accounts for whether previous evaluations of the focal patent were performed by another patent engineer. This variable distinguishes cases where patent engineers previously evaluated the focal patent and should thus already possess information and prior beliefs at the time of evaluation.

Second, in line with previous work (Khanna, Guler, and Nerkar, 2018; Higham, de Rassenfosse, and Jaffe, 2020), I included several patent family level variables that could influence patent engineers' perception of future patent value. *Family size* and *number of claims* can be associated with the scope of patent protection and, thus, with patent value. *Family size* is the number of jurisdictions where patent applications were submitted, while *number of claims* is the maximum number of independent claims across the patents of the family. I counted the maximum number of *forward* and *backward citations* across the family to account for potential signals of quality. The number of granted and abandoned family members at the time of evaluation could have respectively been perceived as positive and negative quality signals by patent engineers. Accordingly, *granted ratio* and *abandoned ratio* are the proportions of respectively granted and abandoned family members to the family size at any time up to the evaluation. I included *filing year* dummy variables to control for variations in patent quality due to time trends and 14 *technology* dummy variables corresponding to one-digit technology classes at Alpha.

Estimation Strategy

I test my hypotheses using Logit models to estimate the likelihood of overestimation and underestimation for my main sample of group evaluations. In robustness tests, I use endogenous treatment-regression models with binary treatment and outcomes to control for potential estimation bias due to self-selection of patent engineers into the decision to rely on group evaluations (Heckman, 1979; Woolridge, 2010). The results indicate no evidence of correlation between the treatment-assignment errors and the outcome errors, that is, the unobservables that may affect the likelihood of over- and underestimation are not correlated with the unobservables that may affect patent engineers' decision to consul other experts. Thus, the coefficient estimates in my main sample are unlikely affected by self-selection or selection on unobservable bias.

In Appendixes B and C, I report results from the endogenous treatment-regression models and from choice models of the selection of which evaluators to involve among the ones available at the time of evaluation. The latter set of models estimate evaluator level factors that likely determined patent engineers' choices, including other evaluators' experience, tenure, expertise depth and breadth, depth and breadth diversity relative to the focal patent engineer, and number of previous evaluations made with the focal patent engineer.

I use a panel dataset with 141 patent engineers (units) and fixed effects models at the patent engineer, patent filing year and technology level. I estimate robust standard errors clustered by patent engineer. Since I am interested in examining the effects on group evaluations of different distributions of expertise among expert evaluators, I included patent engineers who evaluated 15 or more patents. This threshold reduced the patent engineers in my main sample from 161 to 141 but did not exclude any of the other evaluators, indicating that patent engineers tended to consult relatively experienced evaluators (see also Appendix C).

To assess the possibility of multicollinearity, I derive the variance-inflated factors and find that the largest value is 3, which is less than the threshold value of 10. Finally, to facilitate the interpretation and graphical illustration of my moderator hypothesis, I standardize the continuous moderator variables by subtracting the mean and dividing by the standard deviation.

RESULTS

Table 2 reports summary statistics and pair-wise correlations for the variables used in my models. Most of the correlation coefficients are low. I derived variance inflation factors (VIF) for all models (Greene, 2003). All the computed values were less than 3 and the mean VIF was less than 1.6, indicating that multicollinearity is not a concern in the regressions.

--- INSERT TABLE 2 & 3 ABOUT HERE ---

The low correlations between expertise breadth and depth for both patent engineers and the other evaluators confirm that, as discussed above, the two dimensions of expertise are indeed independent. Table 3 reports the frequency counts of expertise breadth and depth for patent engineers and other evaluators for high (above average) and low (below average) values of the two variables. The figures show that the sub-groups of observations are relatively balanced, although the distribution of expertise depth is more skewed towards lower values for the patent engineers with decision authority than for other evaluators. Additionally, Table 2 shows that evaluators tend to have a higher expertise in the focal technology ($mean = 0.56$) than patent engineers ($mean = 0.21$). Indeed, expertise in the technology is one of the key drivers of patent engineers' choice of expert evaluators – see also Appendix C for a more detailed analysis.

In Tables 4, 5 and 6, I report the coefficient estimates of the Logit models. Models 1 and 5 are the baseline models, including only control variables. These models show moderate learning effects as the total experience of both patent engineers and evaluators is negatively associated with both overestimation and underestimation.

Hypothesis 1: Expertise Depth

Hypotheses 1a,b are supported. Hypothesis 1 theorizes that patent engineers with greater expertise in the focal technology are more likely to overestimate and less likely to underestimate its value. To test Hypothesis 1, I introduce patent engineers' expertise depth in Models 2 to 4 (H1a) and Models 6 to 8 (H1b) of Table 4; and evaluators' expertise depth in Models 9 to 11 (H1a) and Models 13 to 15 (H2) of Table 5. Models 12 and 16 in Table 5 are the full models, which include patent engineers' and evaluators' expertise depth and breadth.

The coefficients of the patent engineers' and evaluators' expertise depth variables are positive for overestimation ($B = 0.75$, $p < 0.01$, $B = 0.47$, $p < 0.05$) and negative for underestimation ($B = -1.73$, $p < 0.05$, $B = -0.87$, $p < 0.05$), providing support for Hypothesis 1.

It is worth noting that the coefficients of the patent engineers' and evaluators' expertise breadth variable are not statistically significant in all models, including in the full sample. As I discuss in the robustness tests section, this result was confirmed by alternative operationalizations of expertise breadth.

Table 6 reports effect size estimates for patent engineers' and evaluators' expertise depth and breadth. The results show that the changes in the probabilities of over- and underestimation for a change in depth of expertise are larger for patent engineers than other evaluators. This result was expected since patent engineers have decision authority over evaluations. One standard deviation increase in expertise depth correspond to a 5.2% increase in the probability of overestimation and a 6.3% decrease in the probability of underestimation for patent engineers, and a 2.5% increase in overestimation and a 3.4% decrease in underestimation for other evaluators.

--- INSERT TABLE 6 ABOUT HERE ---

Hypotheses 2 and 3: Depth and Breadth Diversity

Hypotheses 2 and 3 are supported. Hypothesis 2 theorizes that higher group depth diversity is associated with greater evaluation accuracy. Hypothesis 3 theorizes that the positive relationship between depth diversity and evaluation accuracy is negatively moderated by breadth diversity. To test these Hypotheses, I introduce the depth and breadth diversity variables to the baseline Model of evaluation accuracy in Models 18 to 21 of Table 7. The coefficient estimates for depth diversity are positive and significant ($B = 1.57$, $p < 0.01$), providing support to Hypothesis 3. Models 19 and 20 show that the coefficient estimates of breadth diversity are not statistically significant. However, breadth diversity shows a statistically significant interaction with depth diversity. The coefficient estimate for the interaction term is negative and significant ($B = -2.02$, $p < 0.01$), suggesting that breadth diversity negatively moderates the relationship between depth diversity and evaluation accuracy.

I follow established best practice to interpret the significance of the interaction term in nonlinear models and to conduct post hoc analyses of the significance of the moderation effects, since the effect of the interaction between two variables cannot be assessed simply by looking at the sign or statistical significance of the interaction term coefficient (Zelner, 2009; Criscuolo et al 2017). To this end, I use the simulation-based procedure proposed by King, Tomz, and Wittenberg (2000) in the field of political science, which Zelner (2009) has advocated for use in management research. The approach consists of repeatedly drawing estimates from the multivariate normal distribution of the estimated coefficients and the variance matrix through repeated statistical simulation. Using these simulated coefficients, I can derive the change in the predicted probability of the evaluation being accurate, as well as the confidence interval of this change, at two levels of the breadth diversity variable over the entire observed range of depth diversity, while holding all other continuous explanatory variables at their mean.

--- INSERT TABLE 7 & FIGURE 2 ABOUT HERE ---

Figure 2a illustrates the effect of an increase in breadth diversity from its mean value to one standard deviation above the mean on the probability of the evaluation being accurate. In line with my third hypothesis, an increase in breadth diversity reduces the positive effect of depth diversity on evaluation accuracy. Specifically, the slope of the curve becomes negative for higher values of breadth diversity. The curves for low and high breadth diversity cross over near the mean value of depth diversity, suggesting that higher breadth diversity improves evaluation accuracy at low values of depth diversity. This is confirmed in Figure 2b, which reports the difference in the predicted probability of the evaluation being accurate and the 90% confidence interval associated with one standard deviation increase in breadth diversity from its mean value. The interaction effect of breadth diversity is statistically significant across for low ($depth\,diversity\ <\ 0.2$) and high ($depth\,diversity > 0.57$) values of the depth diversity variable. That is, the confidence interval surrounding the difference in predicted shares does not contain zero for low and high values of the depth diversity variable.

The estimates for the other controls are consistent across specifications and confirm our expectations. The total experience for both patent engineers and other evaluators is negatively associated with is overestimation and underestimation and thus positively associated with evaluation accuracy in all models, indicating a learning effect. The workload allocated to the patent engineers at the time of evaluation is negatively associated with evaluation accuracy, although the coefficient for the workload variable is positive and significant only in models of overestimation but not significant in model of underestimation.

Robustness tests

I perform a number of robustness tests to explore some of these findings and assess the validity of the results using alternative specifications of the explanatory variables. In line with my arguments for Hypothesis 1, I expect patent engineers' expertise depth to be associated with over- and underestimation regardless of whether they made evaluations individually or in groups. I estimate Logit models of overestimation and underestimation for the entire sample of evaluations that patent engineers made both individually and in groups (N=17,883), where I include only variables at the patent engineer and patent level. Due to patent engineers' endogenous decision to evaluate in groups, I also estimate Probit models with endogenous binary treatment effects to control for self-selection and selection on unobservables bias – see

Appendix B for more details. The results confirm the results that I obtained for my main sample. The coefficients of the patent engineers' expertise depth variable are positive and significant for overestimation ($B = 0.69,\ p < 0.01$) and negative and significant for underestimation ($B = -1.73, p < 0.05$), providing additional support for Hypothesis 1.

I argued that depth of expertise has relationships of opposite sign with overestimation and underestimation because it is associated with patent ratings. Specifically, if evaluators with greater expertise assign higher ratings all else being equal, they must also be more likely to overestimate and less likely to underestimate the future outcomes of patents. Thus, I test linear models of patent ratings with patent engineers, evaluators and patent level variable equivalent to Models 9 to 12. In line with my expectations, the coefficients of both patent engineers' and evaluators' expertise depth variables are positively and significantly associated with patent ratings ($B = 0.34, p < 0.01, B = 0.14, p < 0.01$).

Contrary to expertise depth, the breadth of expertise of both patent engineers and other evaluators was not associated with patent ratings and with over- or underestimation in all our models. Further, the variable did not have statistically significant interactions effects with expertise depth. I tested alternative operationalisation of expertise breadth. I counted only technology classes with which experience was accumulated over a time period of 6 or 3 years prior to the evaluation, or for which at least 5 or 10 evaluations were made. In line with prior work, I also used the Herfindahl–Hirschman index to derive a measure of individual breadth (Bunderson and Sutcliffe 2002; Lee and Csaszar 2017). None of these alternative operationalisations showed statistically significant relationships with the dependent variables.

Additionally, I constructed a measure of expertise breadth at the group level by aggregating the experience of all the group members. In line with Taylor and Greve (2006), I counted all the technology classes that patent engineers and evaluators evaluated prior to evaluation, omitting double counts of classes. Further, I used the aggregated distribution of experience across classes to compute the Herfindahl–Hirschman index at the group level. These two variables did not show statistically significant relationships with the dependent variables, providing additional support to my argument that group depth diversity plays an equally if not more important role than breadth diversity in contributing to the accuracy of evaluations.

Finally, one of my main arguments for Hypothesis 1 is that evaluators' subjective judgements are influenced by an agency bias, i.e. by self-interested motives for favouring technologies in which they have expertise and for penalizing others. I expect this bias to be more pronounced for the inventors of the focal patent. Indeed, inventors have even stronger incentives to favour their own inventions over competing patents in the portfolio (Fuchs et al 2019). Thus, I introduce a dummy variable in Model 9 to 16 for whether the focal patent inventors contributed to the evaluations. As expected, the coefficients of variable are positive and significant for overestimation ($B = 0.18, p < 0.05$) and negative and significant for underestimation ($B = -0.3, p < 0.05$).

DISCUSSION AND CONCLUSIONS

I examined the effects of differences among the types of expertise of expert evaluators contributing to group evaluations in the context of the management of the patent portfolio of a Fortune 500 ITC firm. My main results are twofold. First, evaluators' individual level of expertise in the focal technology was associated with higher likelihood of overestimation and lower likelihood of underestimation of the value of the technology. In contrast, evaluators' breadth of expertise was not associated with the likelihood of overestimation or underestimation and thus with evaluation accuracy.

Second, evaluations were more likely accurate for larger differences between evaluators' level of expertise in the focal technology, that is for higher depth diversity. Differences between evaluators' distributions of expertise across technologies, i.e. the group's breadth diversity, was not associated with evaluation accuracy. Breadth diversity, however, negatively moderated the relationship between depth diversity and accuracy. Higher breadth diversity decreased the likelihood of evaluations being accurate at high levels of depth diversity and increased the likelihood of being accurate at low levels of depth diversity.

Scholars of organisations generally attribute the advantages of group decisions to the possibility to aggregate broader and richer knowledge from the unique expertise of group members, that is, to the group's breadth diversity (Brodbeck et al 2007; Csaszar and Eggers, 2013). These results indicate that breadth diversity is not beneficial per se, and that it can even diminish the accuracy of evaluations when evaluators have very different levels of expertise in the focal technology. Breadth diversity can improve evaluation accuracy when all evaluators have either high or low levels of expertise in the focal technology. These findings suggest that research on group decisions in organisations should distinguish and emphasize the role of both dimensions of expertise diversity in contributing to the performance of group evaluations.

These findings have two main implications. First, expanding access to unique and more diverse expert knowledge via group evaluations does not necessarily improve the performance of decisions as generally assumed (cf Stasser and Birchmeier 2003). This is a consequence of the fact that high or low levels of expertise in the technology being evaluated induce biased subjective judgements in favour or against the focal technology. My theory and findings suggest that individual biases of opposite sign cancel out as long as evaluators share common knowledge bases that enable the reciprocal heeding of others' perspectives (Gavetti and Warglien, 2015). On the contrary, evaluators with different distributions of expertise tend to hold different representations of the focal technology that create barriers for the interpretation and integration of others' perspectives. It follows that a broader pool of expertise can have the detrimental effect of deterring the mutual attenuation of individual biases on the aggregate evaluation.

A second counterintuitive implication is that increasing access to more advanced knowledge by aggregating the contributions of a larger number of evaluators with expertise does not necessarily improve the performance of evaluations. On the contrary, evaluators with greater expertise would benefit from the contributions of expert evaluators with less or no expertise in the technology. These are further consequences of the relationship between expertise depth and bias and of the role that group evaluations play in attenuating and reinforcing individual biases.

In fact, my findings imply that expanding access to knowledge and attenuating individual biases are two distinct and interdependent mechanisms by which superior evaluation accuracy can be achieved in groups rather than individually. I argued that group evaluations can add value by exposing evaluators to others' perspectives and by inducing the reciprocal updating of their subjective judgements. Importantly, these findings imply that evaluators may update they judgements regardless of the quality of the knowledge and information possessed by others. Even if evaluators without expertise may not be able to provide any additional valuable information to experts of the focal technology, the less positive perspectives of the former can induce the latter to consider a more balanced common ground between opposite opinions. Put it differently, evaluators with expertise who make decisions individually tend to give importance exclusively to the arguments that support their own biased views and do not have an opportunity to reassess or reconcile their subjective judgements with other perspectives.

This study also contributes to the literature on the role of different types of expertise in organisations and on evaluations under uncertainty by examining the link between accuracy and depth and breadth of expertise. Research has examined how dimensions of expertise, such as individual depth or individual and team breadth, influence outcomes such as evaluation ratings (Boudreau et al., 2016, Li, 2017) and project funding (Criscuolo et al, 2017). However, with the exception of formal and conceptual work (Knudsen and Levinthal, 2007; Csaszar and Eggers, 2013), studies of the actual performance of evaluations in organisational contexts are notably missing from the literature. I contribute in two main ways. First, although the total experience of evaluators is positively related to accuracy as expected, their degree of specialisation in the focal technology has a positive effect in reducing underestimation but also a negative effect in increasing overestimation. The net effect of evaluation accuracy is thus very limited or null as the two effects tend to compensate each other. Second, expertise breadth at both the individual and group level was not associated with over- or underestimation and thus with evaluation accuracy in this setting. Future work could explore the relationship between breadth and accuracy in more detail, for instance in different settings where evaluators are less likely subject to agency biases.

Further, these findings have practical implications for the design of group evaluations and decision processes. First, I show how the distribution of expertise across group members can be chosen to optimize the efficacy of group evaluations against their costs. My theory and results suggest that group accuracy is optimized when depth diversity is high and breadth diversity is low. Thus, firms could assign evaluation tasks to groups of evaluators in a controlled way such that they maintain both similar distributions of expertise across technologies (low breadth diversity) and subgroups of evaluators respectively with high and low expertise in selected technologies (high depth diversity). Indeed, this may be difficult to achieve for all the technologies of the portfolio, thus firms may want to ensure that groups with these characteristics exist for the technologies they want to prioritise.

Second, firms may be differentially interested in avoiding overestimation vis-à-vis underestimation and group distributions of expertise can be chosen to satisfy different objectives in this sense. Specifically, cost sensitive firms may be prefer underestimating rather than overestimating the value of technologies to avoid overinvesting. On the contrary, firms with more slack resources may prefer overinvesting and avoiding missing out on valuable opportunities. This approach is analogous to the design of appropriate loss functions in decision theoretic approaches to statistical inference (Casella and Berger, 2010). That is, the balance between degrees of specialisation in the technologies being evaluated and the level of breadth diversity among evaluators can be adjusted as to minimize the likelihood of one type of errors at the expense of increasing the likelihood of the other.

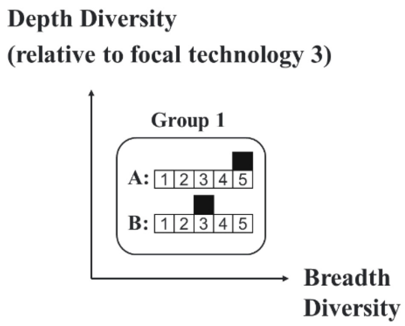
Finally, it is worth noting how the above findings depend on some of the main features of this study's empirical settings. First, technology evaluations were made for a large patent portfolio with the objective to identify opportunities for commercialisation but also to save on maintenance costs. As I discussed, the need to save on costs and to drop technologies from the portfolio likely created incentives for evaluators with expertise to favour the technologies for which their expertise was recognized and valued within the firm. While I would expect the familiarity bias to affect individual preferences across a broad range of contexts, this feature of my setting might exacerbate the influence of the agency bias. In contrast, other settings such as peer review of research grant proposals are less likely to create similar incentives for evaluators (Boudreau et al, 2016, Li, 2017).

Second, my theory and findings depend on the fact that group evaluations were made by a limited number of evaluators who could interact and exchange opinions and recommendations. This study thus extends research on the performance of different aggregation structures (Tindale et al, 2003) and on the wisdom of the crowds phenomenon (Surowiecki, 2004), which generally assume independence of opinions and absence of interactions among group members (Oinas-Kukkonen, 2008). I contribute by examining both effect of interactions among evaluators; and the role that different distribution of knowledge and expertise plays in group evaluations, which, besides a few exceptions (Csaszar and Eggers, 2013), are not discussed or assumed as homogenous in this literature. Future work could further explore the effects of interactions by examining contingencies that may facilitate or compromise exchanges of knowledge among group members. For instance, research could explore the effects of different group hierarchies or communication networks, of relationships of power or conflicts among members, or of the learning of others' competence and biases that may occur after repeated interactions (Christensen and Knudsen, 2010; Nebus, 2006).

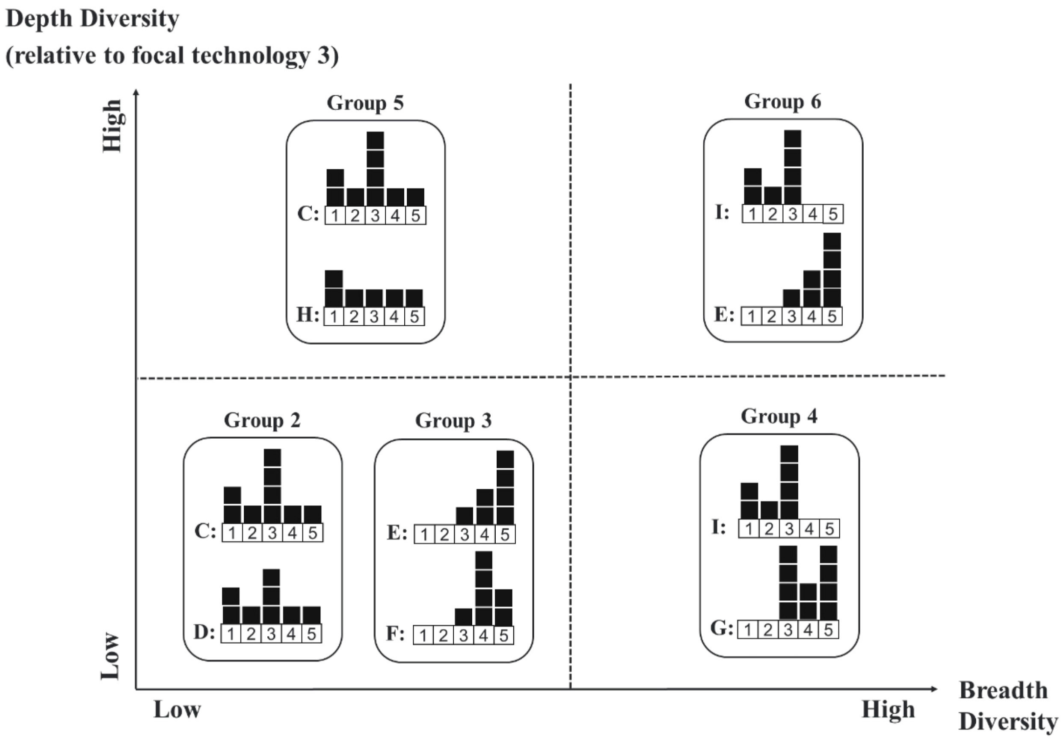
FIGURE 1

Illustration of Evaluators'' Distributions of Expertise and of Depth and Breadth Diversity

(a) Novice Evaluators



(b) Expert Evaluators



*Notes***.** Panel (a) and Panel (b) show illustrative distributions of expertise of novice and expert evaluators, respectively. In this example, eight evaluators identified by the letters A to H are sorted in six two-member groups and accumulated experience across five technology classes. This study does not examine novice evaluators, i.e. evaluators whose total volume of experience is small – see Panel (a). It will instead focus on expert evaluators, i.e. evaluators who accumulated medium to large stocks of experience – see Panel (b). In Panel (b), each group is characterized by high or low values of depth and breadth diversity and represented in the diagram accordingly. Depth diversity is measured relative to the focal technology 3. For instance, Groups 2, 3 and 4 are characterized by low depth diversity because, respectively, evaluators C and D, evaluators E and F, and evaluators I and G have similar experience with technology 3; while Groups 4 and 6 are characterized by high breadth diversity because, respectively, evaluators I and G, and evaluators I and E have similar distributions of experience across technologies.

TABLE 1

Descriptive of Variables

Accurate Evaluation	Binary variable for whether the evaluation correctly estimated future patent commercialization outcomes. It equals 0 when either the overestimation or underestimation variables equal 1.
Breadth Diversity	Average of the cosine distanced between the expertise breadth of each pair of evaluators who contributed to the evaluation.
Depth Diversity	Average of the absolute differences between the expertise depth of each pair of evaluators who contributed to the evaluation.
Evaluators: Experience	Average of the number of evaluations made by expert evaluators and inventors prior to evaluation date, and of the number of inventions disclosed by the inventors prior to evaluation date.
Evaluators: Expertise Depth	Average of the proportion of evaluations of the patent's technology class to total number of evaluations made by expert evaluators and inventors prior to evaluation date, and of the proportion of inventions disclosed for the patent's technology class to total number of inventions disclosed by the inventors prior to evaluation date.
Evaluators: Expertise Breadth	Average of the number of technology classes that expert evaluators and inventors evaluated 5 or more time over the 3 years prior to evaluation date, and of the number of technology classes in which the inventors disclosed inventions over the 3 years prior to evaluation date.
Evaluators: Opinion Requested	Binary variable for whether evaluators' opinions were requested by the patent engineer and obtained at the time of evaluation.
Evaluators: Total Available Opinions	Number of opinions given by evaluators for the patent prior to evaluation date
Evaluators: Pool size	Number of experts whom patent engineers can request opinions to at the time of evaluation
Inventor: Team Size	Number of patent's inventors
Inventor: Team Tenure	Average number of years since the year of the first invention disclosure of each inventor on the team at the time of evaluation
Inventor: Opinion Requested	Binary variable for whether inventors' opinions were requested by the patent engineer and obtained at the time of evaluation
Inventor & Expert: Opinion Requested	Binary variable for whether both experts' and inventors' opinions were requested by the patent engineer and obtained at the time of evaluation
Inventor: Prior Opinions to Patent Engineer	Average across the inventors team of number of opinions requested and obtained by the patent engineer for the focal and any other patent prior to evaluation date

Overestimation	Binary variable, indicates whether a Type I error was committed, i.e. whether the patent engineer maintained or increased the patent's rating, and the patent was not commercialized by the end of its life
Patent: Abandoned	Proportion of abandoned to total number of patent applications filed for the patent family at the time of evaluation
Patent: Age	Number of years since filing at the time of evaluation
Patent: Backward/Forward Citations	Number of patent's backward/forward citations
Patent: Claims	Number of patent claims
Patent: Flagged	Binary variable for whether the patent has been identified as having potential for standardization or implementation prior to evaluation date
Patent: Granted	Proportion of granted to total number of patent applications filed for the patent family at the time of evaluation
Patent: Transferred Case	Binary variable that equals 1 if the patent has been evaluated by other patent engineers at the time of evaluation
Patent: Uncertainty	Binary variable for whether the patent's rating prior to evaluation date is 2 or 3 (out of 5)
Patent Engineer: Experience	Number of evaluations made by the patent engineer prior to evaluation date
Patent Engineer: Expertise Breadth	Number of technology classes that the patent engineer evaluated 5 or more time over the 3 years prior to evaluation date.
Patent Engineer: Expertise Depth	Proportion of evaluations of the patent's technology class to total number of evaluations made by the patent engineer prior to evaluation date
Patent Engineer: Workload	Number of evaluations made by the patent engineer during the focal week, including the focal evaluation.
Underestimation	Binary variable, indicates whether a Type II error was committed, i.e. whether the patent engineer decreased the patent's rating, and the patent was successfully commercialized by the end of its life

TABLE 2
Descriptive Statistics and Correlation Matrix (N=3,819)

	VARIABLE	MIN	MAX	MEAN	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
1	Overestimation	0	1	0.24										
2	Underestimation	0	1	0.14	-0.2*									
3	Patent Engineer: Experience	15	578	115.01	0.07*	-0.03*								
4	Evaluators: Experience	23	742	75.86	0	-0.01	0.13*							
5	Patent Engineer: Expertise Depth	0	1	0.21	0	-0.01	0	0.06*						
6	Patent Engineer: Expertise Breadth	1	102	52.33	0.08*	-0.06*	0.39*	0.01	-0.27*					
7	Evaluators: Expertise Depth	0	1	0.56	-0.02*	-0.01	-0.07*	-0.19*	0.26*	-0.1*				
8	Evaluators: Expertise Breadth	1	86	20.4	0.03*	-0.02*	0.19*	0.39*	0.04*	0.14*	-0.35*			
9	Depth Diversity	0	272	15.26	0.07*	-0.05*	0.18*	0.05*	0.32*	0.03*	0	-0.02*		
10	Breadth Diversity	0	0.99	0.4	0.05*	-0.01	-0.02*	-0.18*	-0.59*	0.13*	-0.11*	-0.17*	-0.16*	
11	Inventors' Opinions Requested	0	1	0.78	0.0	0	-0.07*	0.02*	0.01	-0.05*	0.27*	0.1*	-0.03*	-0.08*
12	Patent: Transferred Case	0	1	0.39	0.04*	-0.03*	0.19*	0.11*	0.01	0.08*	-0.04*	0.16*	0	0.02*
13	Patent: Uncertainty	0	1	0.85	0.02*	0.03*	0.02*	0.01	0.01	0.04*	-0.01	0.01	0.04*	-0.01
14	Patent Engineer: Workload	1	58	2.84	-0.04*	0.02*	0	-0.06*	0	-0.02*	0.01	-0.1*	0	0
15	Patent: Granted	0	1	0.12	-0.02*	0.07*	0.13*	-0.1*	-0.08*	0.02*	-0.04*	-0.1*	0.01	0.12*
16	Patent: Claims	1	82	21.47	0.05*	-0.03*	0.05*	0.09*	0.04*	0.05*	-0.03*	0.14*	0.05*	-0.03*
17	Patent: Forward Citations	0	477	31.57	0.02*	0.01	0.04*	0.06*	0.02*	0.06*	-0.05*	0.17*	-0.03*	0.01
18	Patent: Backward Citations	0	249	14.75	0.07*	-0.02	0.07*	0.05*	0.01	0.09*	-0.06*	0.15*	0.01	0.02*
19	Patent: Age	0	17	4.63	0	0.04*	0.27*	-0.03*	-0.07*	0.12*	-0.09*	0	0.02*	0.13*
20	Patent: Abandoned	0	1	0.12	-0.02*	0.02*	0.19*	-0.02	-0.02*	0.09*	-0.03*	0.01	0.01	0.05*

TABLE 2
(Continued)

	VARIABLE	MIN	MAX	MEAN	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)
12	Patent: Transferred Case	0	1	0.39	-0.02*								
13	Patent: Uncertainty	0	1	0.85	0.07*	-0.08*							
14	Patent Engineer: Workload	1	58	2.84	-0.11*	-0.07*	-0.06*						
15	Patent: Granted	0	1	0.12	-0.13*	0.14*	-0.14*	0.14*					
16	Patent: Claims	1	82	21.47	0.02*	0.04*	0.02*	-0.08*	-0.16*				
17	Patent: Forward Citations	0	477	31.57	-0.05*	0.05*	-0.06*	-0.03*	0.07*	0.15*			
18	Patent: Backward Citations	0	249	14.75	-0.03*	0.05*	-0.03*	-0.03*	-0.05*	0.15*	0.32*		
19	Patent: Age	0	17	4.63	-0.15*	0.32*	-0.16*	0.1*	0.72*	-0.15*	0.09*	-0.01	
20	Patent: Abandoned	0	1	0.12	-0.06*	0.23*	-0.06*	0.05*	0.28*	-0.08*	0	-0.02*	0.51*

TABLE 3
Frequency Counts of Expertise Depth and Breadth for Patent Engineers and Evaluators

Patent Engineer: Expertise Depth	Patent Engineer: Expertise Breadth			Evaluators: Expertise Depth	Evaluators: Expertise Breadth		
	Low (<mean)	High (>mean)	Total		Low (<mean)	High (>mean)	Total
Low (<mean)	1,636	1,518	3,154	Low (<mean)	1,122	984	2,106
	52	48	100		53.28	46.72	100
	79.57	86.1	82.59		54.92	55.41	55.15
High (>mean)	420	245	665	High (>mean)	921	792	1,713
	63	37	100		53.77	46.23	100
	20.43	13.9	17.41		45.08	45.59	44.85
Total	2,056	1,763	3,819	Total	2,043	1,776	3,819
	53.84	46.16	100		53.5	46.5	100
	100	100	100		100	100	100

TABLE 4
Logit Models on Overestimation and Underestimation (N=3,819)

	Overestimation (Type I)				Underestimation (Type II)			
VARIABLES	1	2	3	4	5	6	7	8
Hypothesis 1: Patent Engineer: Expertise Depth		0.62**		0.70**		-1.94**		-2.19**
		(0.24)		(0.28)		(0.95)		(0.97)
Patent Engineer: Expertise Breadth			0.08	0.29			-0.21	-0.45
			(0.18)	(0.21)			(0.32)	(0.35)
Patent Engineer: Experience	-0.06***	-0.06***	-0.06***	-0.06***	-0.01***	-0.01***	-0.01***	-0.01***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Expert: Experience	-0.00***	-0.00***	-0.00***	-0.00***	-0.01*	-0.01*	-0.01*	-0.01*
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Patent: Transferred Case	0.05	0.05	0.06	0.06	-0.41**	-0.41**	-0.41**	-0.42**
	(0.09)	(0.09)	(0.09)	(0.09)	(0.19)	(0.19)	(0.19)	(0.19)
Patent: Prior Evaluations (#)	0.03	0.03	0.04	0.03	-0.04	-0.04	-0.05	-0.05
	(0.03)	(0.03)	(0.03)	(0.03)	(0.07)	(0.07)	(0.07)	(0.07)
Patent: Uncertainty	0.04	0.04	0.04	0.04	0.29	0.29	0.29	0.28
	(0.10)	(0.10)	(0.10)	(0.10)	(0.20)	(0.20)	(0.20)	(0.20)
Patent Engineer: Workload	0.02**	0.02**	0.02**	0.02**	0.01	0.01	0.01	0.01
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Patent: Granted	0.35*	0.33*	0.34*	0.31	0.28	0.28	0.19	0.32
	(0.20)	(0.20)	(0.20)	(0.20)	(0.36)	(0.36)	(0.36)	(0.37)
Patent: Forward Citations	-0.00	-0.00	-0.00	-0.00	0.00	0.00	0.00	0.00
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)

Patent: Backward Citations	0.01***	0.01***	0.01***	0.01***	0.01**	0.01**	0.01**	0.01**
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)

TABLE 4
(Continued)

	Overestimation (Type I)				Underestimation (Type II)			
VARIABLES	1	2	3	4	5	6	7	8
Inventors' Opinions Requested	0.04	0.04	0.05	0.04	-0.02	-0.01	-0.02	-0.01
	(0.09)	(0.09)	(0.09)	(0.09)	(0.18)	(0.19)	(0.19)	(0.19)
Patent: Abandoned	-0.53**	-0.54**	-0.52**	-0.54**	0.29	0.26	0.29	0.27
	(0.25)	(0.25)	(0.25)	(0.25)	(0.48)	(0.48)	(0.48)	(0.48)
Constant	-1.97***	-1.91***	-1.86**	-2.07***	-1.95**	-1.89**	-1.78**	-1.73**
	(0.74)	(0.74)	(0.74)	(0.74)	(0.74)	(0.74)	(0.75)	(0.75)
Patent Engineer Dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Filing Year Dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Technology Dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Robust standard error clustered by patent engineer.
*p<0.1, **p<0.05, ***p<0.01

TABLE 5
Logit Models on Overestimation and Underestimation (N=3,819)

	Overestimation (Type I)				Underestimation (Type II)			
VARIABLES	9	10	11	12	13	14	15	16
Hypothesis 1: Patent Engineer: Expertise Depth				0.75***				-1.73**
				(0.29)				(1.00)
Patent Engineer: Expertise Breadth				0.33				-0.40
				(0.21)				(0.35)
Patent Engineer: Experience	-0.06***	-0.06***	-0.06***	-0.06***	-0.01***	-0.01***	-0.01***	-0.01***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Hypothesis 1: Evaluators: Expertise Depth	0.41*		0.43**	0.47**	-0.99***		-0.96***	-0.87**
	(0.29)		(0.23)	(0.23)	(0.38)		(0.51)	(0.51)
Evaluators: Expertise Breadth		0.81	0.69	1.11		1.04	0.01	-0.19
		(0.27)	(0.21)	(0.36)		(0.57)	(0.76)	(0.77)
Evaluator: Experience	-0.00***	-0.00***	-0.00***	-0.00***	-0.01*	-0.01*	-0.01*	-0.01*
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Patent: Transferred Case	0.07	0.06	0.07	0.07	-0.44**	-0.42**	-0.43**	-0.43**
	(0.09)	(0.09)	(0.09)	(0.09)	(0.19)	(0.19)	(0.19)	(0.19)
Patent: Prior Evaluations (#)	0.04	0.04	0.04	0.04	-0.05	-0.05	-0.05	-0.05
	(0.03)	(0.03)	(0.03)	(0.03)	(0.07)	(0.07)	(0.07)	(0.07)
Patent: Uncertainty	-0.05	-0.05	-0.06	-0.06	0.29	0.28	0.28	0.27
	(0.10)	(0.10)	(0.10)	(0.10)	(0.20)	(0.20)	(0.20)	(0.20)
Patent Engineer: Workload	0.02***	0.02**	0.02**	0.02**	0.01	0.01	0.01	0.01
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)

Patent: Granted	0.35*	0.38*	0.39**	0.39**	0.19	0.21	0.19	0.31
	(0.20)	(0.20)	(0.20)	(0.20)	(0.36)	(0.36)	(0.36)	(0.37)

TABLE 5
(Continued)

	Overestimation (Type I)				Underestimation (Type II)			
VARIABLES	9	10	11	12	13	14	15	16
Patent: Forward Citations	0.00	0.00	0.00	0.00	-0.00	-0.00	-0.00	-0.00
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Patent: Backward Citations	0.00	-0.00	0.00	-0.00	0.00	0.00	0.00	0.00
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Inventors' Opinions Requested	0.04	0.04	0.05	0.04	-0.02	-0.01	-0.02	-0.01
	(0.09)	(0.09)	(0.09)	(0.09)	(0.18)	(0.19)	(0.19)	(0.19)
Patent: Abandoned	-0.48*	-0.49**	-0.48**	-0.47**	0.31	0.31	0.31	0.30
	(0.25)	(0.25)	(0.25)	(0.25)	(0.48)	(0.48)	(0.48)	(0.48)
Constant	-1.83**	-1.87**	-2.45***	-3.11***	-1.54**	-2.60***	-1.56***	-1.32***
	(0.73)	(0.74)	(0.77)	(0.83)	(0.74)	(0.85)	(0.99)	(1.00)
Patent Engineer Dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Filing Year Dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Technology Dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Robust standard error clustered by patent engineer.
*p<0.1, **p<0.05, ***p<0.01

TABLE 6
Effect sizes

xtlogit:	Changes in Pr(y) Number of obs = 3819			
Expression:	Pr(Overestimation=1), predict(pr)		Pr(Underestimation=1), predict(pr)	
	Change	p-value	Change	p-value
Expert: Depth Expertise				
+1	0.173	0.015	-0.210	0.031
+1SD	0.025	0.015	-0.034	0.047
Marginal	0.165	0.014	-0.192	0.052
Patent Engineer: Breadth Expertise				
+1	0.086	0.179	0.073	0.513
+1SD	0.017	0.172	0.012	0.457
Marginal	0.083	0.168	0.033	0.445
Evaluators: Breadth Expertise				
+1	0.092	0.097	-0.032	0.482
+1SD	0.020	0.092	-0.001	0.612
Marginal	0.089	0.089	-0.046	0.624
Patent Engineer: Depth Expertise				
+1	0.226	0.006	-0.255	0.000
+1SD	0.052	0.008	-0.063	0.008
Marginal	0.216	0.008	-0.248	0.010
Average prediction	0.373		0.164	

TABLE 7
Logit Models on Accurate Evaluation (N=3,819)

	Accurate Evaluation				
VARIABLES	17	18	19	20	21
H2: Depth Diversity		1.47***		1.57***	1.74***
		(0.39)		(0.41)	(0.59)
Breadth Diversity			-0.26	-0.26	0.59**
			(0.14)	(0.14)	(0.17)
H3: Depth Diversity # Breadth Diversity					-2.02***
					(0.68)
Evaluator: Experience	0.00**	0.01**	0.01**	0.01**	0.01**
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Patent Engineer: Experience	0.06***	0.06***	0.06***	0.01***	0.01***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Inventor: Opinion Requested	-0.04	-0.04	-0.05	-0.05	-0.05
	(0.09)	(0.09)	(0.09)	(0.09)	(0.09)
Patent: Transferred Case	0.11	0.11	0.11	0.11	0.11
	(0.10)	(0.10)	(0.10)	(0.10)	(0.10)
Patent: Abandoned	0.40*	0.40	0.40	0.39	0.38
	(0.24)	(0.24)	(0.24)	(0.24)	(0.24)
Patent: Uncertainty	-0.06*	-0.06*	-0.05*	-0.05	-0.05
	(0.10)	(0.10)	(0.10)	(0.10)	(0.10)
Patent Engineer: Workload	-0.02**	-0.02**	-0.02**	-0.02**	-0.02**
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)

TABLE 7
(Continued)

	Accurate Evaluation				
VARIABLES	17	18	19	20	21
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Patent: Granted	-0.39*	-0.39*	-0.39*	-0.39*	-0.38
	(0.23)	(0.23)	(0.23)	(0.23)	(0.23)
Patent: Abandoned	0.40*	0.40*	0.40*	0.39	0.38
	(0.24)	(0.24)	(0.24)	(0.24)	(0.24)
Patent: Claims	-0.00	-0.00	-0.00	-0.00	-0.00
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Patent: Forward Citations	-0.00	-0.00	-0.00	-0.00	-0.00
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Patent: Backward Citations	-0.01***	-0.01***	-0.01***	-0.01***	-0.01***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Constant	-1.80**	-2.04***	-1.15	-1.67**	-1.52**
	(0.74)	(0.74)	(1.09)	(0.74)	(0.77)
Patent Engineer Dummies	Yes	Yes	Yes	Yes	Yes
Filing Year Dummies	Yes	Yes	Yes	Yes	Yes
Technology Dummies	Yes	Yes	Yes	Yes	Yes

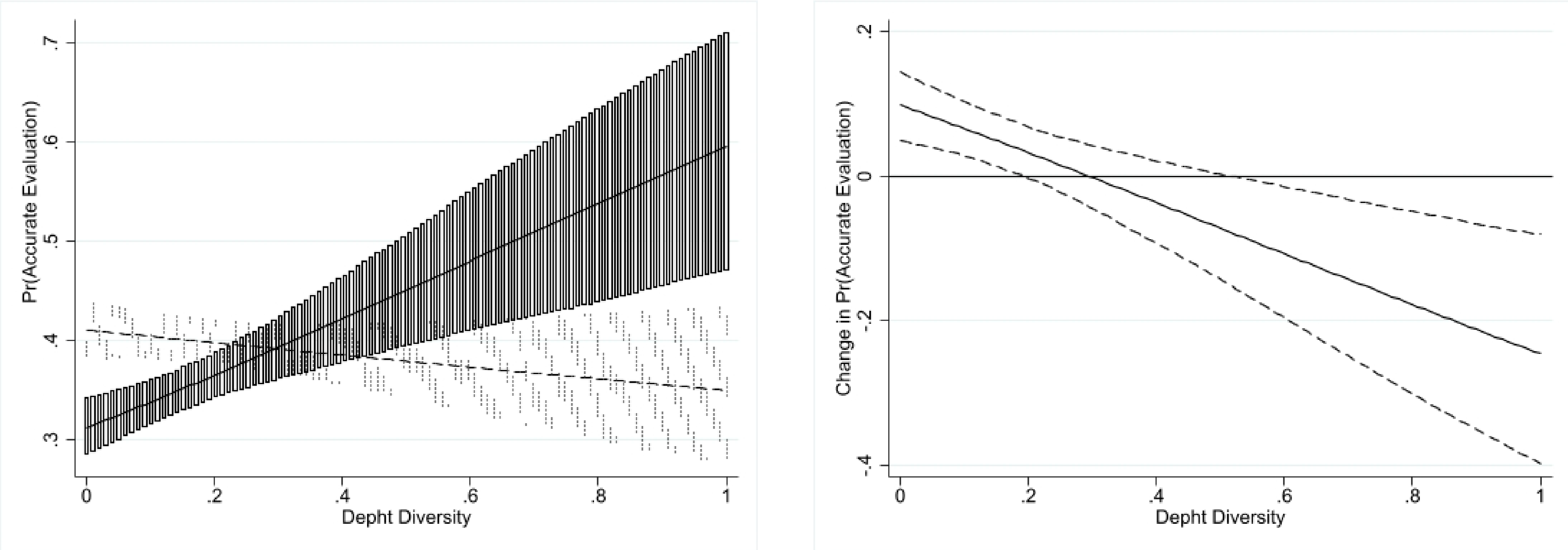
Notes: Robust standard error clustered by patent engineer.

*p<0.1, **p<0.05, ***p<0.01

FIGURE 2
A and B: Moderation Effect of Breadth Diversity on Depth Diversity

A Predicted probability of the evaluation being accurate for low (—) and high (---) breadth diversity

B Delta predicted probability of the evaluation being accurate for low vs and high breadth diversity

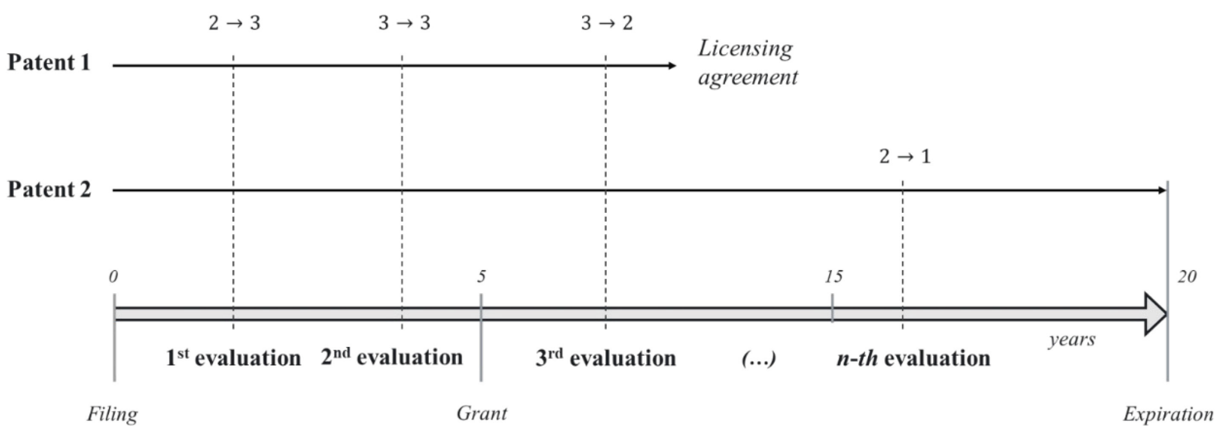


Notes: These graphs show the moderating effect of breadth diversity associated with an increased level of breadth diversity from its mean value (dashed line) to one standard deviation above the mean (continuous line). I obtained the graphs using coefficient estimates from Model 21 of Table 7, considering an evaluation with mean values for all other continuous variables. Dashed lines in Figure 2B show 90% confidence intervals for the difference in the predicted probability of the evaluation being accurate.

APPENDIX A
TABLE 1A
Definition of *Overestimation, Underestimation*, and *Accurate Evaluation*

Patent engineer's decision (*ex-ante*)		Patent outcome (*ex-post*)	
		A. Positive outcome	*B. Negative outcome*
Rating confirmed or increased	1. *Positive prospect*	**Accurate**	**Overestimation** (Type I Error)
Rating decreased	*2. Negative prospect*	**Underestimation** (Type II Error)	**Accurate**

FIGURE 1A
Example of how I compute *Overestimation, Underestimation*, and *Accurate Evaluation.*



		1st evaluation	2nd evaluation	3rd evaluation	*n-th* evaluation
Group 1	Rating change (ΔR)	1	0	-1	
	Overestimation	**0**	**0**	**0**	
	Underestimation	**0**	**0**	**1**	
	Accurate	**1**	**1**	**0**	
Group 2	Rating change (ΔR)	1	0	-1	-1
	Overestimation	**-1**	**-1**	**0**	**0**
	Underestimation	**0**	**0**	**0**	**0**
	Accurate	**0**	**0**	**1**	**1**

In this example, patent engineers 1 and 2 are responsible for periodic re-evaluations of Patents 1 and 2, respectively. The timeline shows that both patents were granted 5 years after filing and that the first two evaluations were made while patents were still under prosecution. Patent 1 was successfully licensed to a third party after grant, while Patent 2 expired at the end of its legal life without being commercialized. That is, the eventual outcomes for Patents 1 and 2 were positive and negative, respectively. For illustrative purposes, both patent engineers evaluated their respective patents at equal time periods after filing and always involved other evaluators. Groups did not necessarily involve the same evaluators when patents were re-evaluated, but both patents were assigned the same ratings by the respective groups. Both ratings increased from 2 to 3 at the first evaluation, were confirmed at 3 at the second evaluation, and decreased to 2 at the third evaluation. Patent 1 was removed from the first patent engineer's list of assigned patents because it was commercialized, and additional evaluations were no longer necessary. Patent 2, instead, was periodically re-evaluated until its expiration date. The table shows the values taken by the dummy variables as the ratings were changed or confirmed. For instance, the rating decrease decided at the third evaluation corresponds to underestimation for Group 1 and to an accurate evaluation for Group 2.

APPENDIX B
Table 1B
Probit Models with Endogenous Binary Treatment Effects (N=17,883)

	Group Evaluation	Overestimation (Type I)			Underestimation (Type II)		
VARIABLES	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7
Patent Engineer: Expertise Depth	-0.24*	0.62**		0.70**	-1.94**		-2.19**
	(0.14)	(0.24)		(0.28)	(0.95)		(0.97)
Patent Engineer: Expertise Breadth	0.00		0.08	0.29		-0.21	-0.45
	(0.00)		(0.18)	(0.21)		(0.32)	(0.35)
Patent Engineer: Experience	-0.01	-0.06***	-0.06***	-0.06***	-0.01***	-0.01***	-0.01***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Expert: Experience	0.00	-0.00***	-0.00***	-0.00***	-0.01*	-0.01*	-0.01*
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Inventor: Team Size	0.03***	0.00	0.00	0.00	0.04	0.04	0.04
	(0.01)	(0.00)	(0.00)	(0.00)	(0.23)	(0.23)	(0.23)
Expert: Available Pool Size	0.03*	0.31	0.31	0.31	0.52	0.52	0.52
	(0.02)	(0.01)	(0.01)	(0.01)	(0.03)	(0.03)	(0.03)
Patent: Transferred Case	0.10***	0.05	0.06	0.06	-0.41**	-0.41**	-0.42**
	(0.04)	(0.09)	(0.09)	(0.09)	(0.19)	(0.19)	(0.19)
Patent: Prior Evaluations (#)	-0.06***	0.03	0.04	0.03	-0.04	-0.05	-0.05
	(0.02)	(0.03)	(0.03)	(0.03)	(0.07)	(0.07)	(0.07)
Patent: Uncertainty	0.11***	0.04	0.04	0.04	0.29	0.29	0.28
	(0.04)	(0.10)	(0.10)	(0.10)	(0.20)	(0.20)	(0.20)
Patent Engineer: Workload	-0.02***	0.02**	0.02**	0.02**	0.01	0.01	0.01
	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)

Patent: Granted	-0.06	0.33*	0.34*	0.31	0.28	0.19	0.32
	(0.06)	(0.20)	(0.20)	(0.20)	(0.36)	(0.36)	(0.37)
Patent: Forward Citations	-0.00	-0.00	-0.00	-0.00	0.00	0.00	0.00
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Patent: Backward Citations	0.00	0.01***	0.01***	0.01***	0.01**	0.01**	0.01**
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Expert Opinion Requested		0.30	0.30	0.30	-0.05	-0.05	-0.05
		(0.09)	(0.09)	(0.09)	(0.13)	(0.13)	(0.13)
$\boldsymbol{\rho}$	-0.94						
	(0.06)						
$\boldsymbol{\sigma}$	-1.31						
	(0.04)						
Constant	0.30	-1.91***	-1.86**	-2.07***	-1.89**	-1.78**	-1.73**
	(0.19)	(0.74)	(0.74)	(0.74)	(0.74)	(0.75)	(0.75)
Patent Engineer Dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Filing Year Dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Technology Dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Robust standard error clustered by patent engineer.
*p<0.1, **p<0.05, ***p<0.01

Patent engineers have decision authority over evaluations and over the decision to evaluate individually or in groups. The decision to consult other evaluators is likely not random and may be influenced by individual, evaluation- or patent-level characteristics. These include unobservable variables that may simultaneously influence both the decision to consult other evaluators and evaluation accuracy. This self-selection and selection on unobservables could lead to biased estimates if not properly accounted for, as it violates the assumption of exogeneity necessary for ordinary least squares regression.

To address this issue, I estimate endogenous treatment-regression models with binary treatment and outcomes that include two instrumental variables. These use a Probit model for the outcome and a normal distribution to model the deviation from the conditional independence assumption imposed by the estimators (Heckman, 1979; Woolridge, 2010). Coefficient estimates are obtained for two sets of regression equations, namely the treatment equation and the outcome equation. In this setting, the treatment equation provides coefficient estimates of the treatment variable *group evaluation*, while the two outcome equations provide coefficient estimates of overestimation and underestimation.

My chosen instrumental variables are the number of other patent engineers and technology experts available to contribute to the evaluations and the size of the patent inventor team. Both variables are likely related to patent engineers' choice as to whether to consult other evaluators and unrelated to the outcomes of evaluations. I measure the number of available evaluators by counting the number of evaluators who contributed to group evaluations and who were nominated by inventors and patent boards to contribute to patenting decisions during the three months prior to evaluation date. I compute the same variable for time periods of 1 and 6 months as alternative operationalisations. There were a minimum of 6 and up to 279 available evaluators at any time, and 103 on average (SD=76). Variations of the number of available evaluators were seasons (e.g. due to holyday periods) or due to periods of heavier workloads at Alpha. The second instrument, the focal patent's inventor team size, ranged between 1 and 8 (88 teams

(<1%) had up to 16 inventors), for an average of 2.14 (SD=1.4). Patent engineers often did not consult the entire team but only chose one or a few of them.

Table 1B shows the coefficient estimates for the treatment equation, i.e. for the variable *Group Evaluation*, and for the outcome equations, i.e. for the variables *Overestimation* and *Underestimation*. The main result is that unobservables that may affect the decision to consult other evaluators are not associated with unobservables that affect overestimation or underestimation. Specifically, the likelihood-ratio test indicates that I cannot reject the null hypothesis of no correlation between the treatment-assignment errors and the outcome errors. The estimated correlation between the treatment-assignment errors and the outcome errors, ρ, is not statistically significant. Thus, the coefficient estimates in my main sample are unlikely affected by self-selection or selection on unobservable bias.

The other estimates offer interesting insights on the determinants of patent engineers' decision to rely on group evaluations rather than on their judgement. All else being equal, it is reasonable to expect that patent engineers with more experience or greater expertise in the focal technology are more likely to rely on their own judgement. The coefficient for the patent engineers' experience variable was negative but not statistically significant, while the coefficient for the patent engineers' expertise depth is negative and statistically significant ($B = -0.01, p < 0.01$). This result provides indicative support to my assumption that evaluators with expertise tend to develop (over)confidence in their own judgement especially in the technologies in which they have greater expertise. Further, the coefficient for the patent engineers' experience variable coefficient is negative and highly significant in additional unreported models where I estimate the decision to consult only patent inventors ($B = -0.01, p < 0.01$). This result may suggest that patent engineers gradually learned that inventors were overly optimistic about their own inventions. Finally, patent engineers' expertise breadth was not associated with the decision to consult other experts.

The instrumental and control variables are in line with my expectations. The coefficient estimates of *inventor team size* and *expert available pool size* are positive and significant, while they were not significant in all models of overestimation and underestimation. The likelihood of relying on group evaluations was higher in conditions of greater uncertainty. The coefficient estimate of *patent uncertainty* is positive and highly significant. Similarly, I expect that evaluations were more uncertain when patent engineers had less information from previous evaluations of the focal patent. Patent engineers had less information about patents that were originally evaluated by other patent engineers and that were subsequently transferred to them (*patent: transferred case*); and that they had more information the higher the number of previous re-evaluations of the focal patent (*Patent: Prior Evaluations (#)*). Accordingly, the coefficient estimates of these variables are significant and respectively positive and negative. Finally, the coefficient estimate of the patent engineers' workload variable is negative and highly significant. This is in line with one the premises of this study that group evaluations are costly and that it is important to examine the conditions under which they are more or less likely to improve evaluation accuracy.

APPENDIX C

Table 1C

Rare-event Logit Models of Patent Engineers' Choice of Evaluators (N=42,015)

VARIABLES	1	2	3	4	5
Evaluators: Depth Expertise		1.90***	2.64***		
		(0.06)	(0.06)		
Evaluators: Breadth Expertise			-0.05***		
			(0.00)		
Patent Engineer – Evaluator Depth Diversity				1.23	1.47
				(0.05)	(0.06)
Patent Engineer – Evaluator Breadth Diversity					-1.80***
					(0.08)
Inventor: Consulted	-0.36***	-0.94***	-0.69***	-0.93***	-0.89***
	(0.05)	(0.06)	(0.06)	(0.06)	(0.06)
Evaluators: Experience	-0.01***	-0.01***	-0.01***	-0.01***	-0.01***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Patent Engineer: Prior Interactions with Evaluator	0.65***	0.60***	0.59***	0.60***	0.52***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Constant	-63.64***	-36.68***	-74.91***	-37.77***	-60.78***
	(3.48)	(3.62)	(2.28)	(1.49)	(3.47)

I am interested in obtaining an econometric model of patent engineers' choices of which evaluators to involve in group evaluations among a relatively high number of available expert evaluators at Alpha. I follow the estimation strategy used by Haas, Criscuolo and George (2015).

The unit of analysis is a patent engineer-evaluator dyad. I constructed a matrix of all patent engineer-by-evaluator dyads in which the *i,j* cell is 1 if patent engineer *i* consulted evaluator *j* (realized dyad) or 0 otherwise (non-realized dyad). The patent engineers in these dyads included all the 140 individuals who chose at least one evaluator for group evaluations during the observation period. I defined the risk set of evaluators to include all possible evaluators that were available to be consulted at the time of evaluation. That is, I measure the number of available evaluators by counting the number of evaluators who contributed to group evaluations and who were nominated by inventors and patent boards to contribute to patenting decisions during the three months prior to evaluation date. I compute the same variable for time periods of 1 and 6 months as alternative operationalisations. There were a minimum of 6 and up to 279 available evaluators at any time, and 103 on average (SD=76). Variations of the number of available evaluators were seasonal (e.g. due to holyday periods) or due to periods of heavier workloads at Alpha.

The resulting dataset consisted of 542,145 possible patent engineer-evaluator dyads, of which 1,742 were coded 1 (realized dyads) and 540,403 were coded 0 (non-realized dyads). Constructing the dataset in this way enabled me to compare realized dyads to non-realized dyads, following the analytic approach taken in previous studies of tie formation between firms (e.g. Gulati, 1995). However, the dataset was characterized by a preponderance of zeros due to the large number of non-realized dyads. The analysis of a dataset with very few positive events (less than 1%) cannot be undertaken using a standard logit model because it will underestimate the probability of a positive outcome (i.e., a match between a patent engineer and an evaluator) (King and Zeng, 2001). The dataset was also characterized by non-independence in the error terms arising from the fact that both patent engineers and evaluators could appear many times in the dataset. This issue of network autocorrelation could lead to underestimation of standard errors (Krackhardt, 1988). To address these concerns, I followed previous studies of tie formation in sparse networks (e.g. Hallen, 2008) by using a choice-based sampling technique and testing our hypotheses using a rare-event Logit model. The choice-based sampling technique included all the realized dyads and a randomly extracted sample of corresponding non-realized dyads. For each realized dyad in which patent engineer *i* consulted evaluator *j*, I randomly selected 10 non-realized dyads from the sample of evaluators whom patent engineer *i* could have consulted but did not (i.e., those among the available evaluators prior to evaluation date). While this choice-based sampling technique resolves concerns created by a preponderance of zeros in the dataset, it can bias the logit estimates because the proportion of positive outcomes in the sample is different from that in the underlying population of potential dyads. To correct this bias, I used weighted exogenous sampling maximum-likelihood estimation (WESMLE), an approach that weights the contribution of each dyad to the likelihood function and is better than alternative approaches for large samples (King and Zeng, 2001). Additionally, I clustered the standard errors on the patent engineer (Hallen, 2008), since each patent engineer appears in one realized dyad and ten unrealized dyads (i.e., the patent engineer is constant across eleven observations). I used Tomz's (2003) ReLogit Stata procedure to estimate the Logit models. Finally, I utilized the longitudinal nature of the dataset by constructing the explanatory and control variables to minimize reverse causality by measuring them in the period prior to the focal match/non-match.

The results are in line with my expectations. The coefficients of the evaluators' expertise depth variable are positive and significant ($B = 2.64, p < 0.01$). As expected, patent engineers consulted evaluators with greater expertise in the focal technology. Relatedly, the absolute difference between patent engineers' and evaluators' expertise in the focal technology – i.e. *Patent Engineer – Evaluator Depth Diversity* was not associated with evaluator choice. In additional underreported analyses, I computed *Patent Engineer – Evaluator Depth Diversity* as the raw difference (as opposed to the absolute value of the difference) between evaluators' and patent engineers' expertise, which resulted positively and significantly associated with evaluator choice. That is, patent engineers tended to consult other evaluators with greater expertise than their own in the focal technology.

Interestingly, patent engineers also preferred evaluators with similar distributions of expertise relative to their own. The coefficients of the *Patent Engineer – Evaluator Breadth Diversity* variable is negative and significant ($B = -1.80, p < 0.01$). This provides additional support to one of my main arguments that evaluators are more likely to integrate and heed the opinions of similar others. Relatedly, patent engineers were more likely to consult evaluators with whom they had a higher number of previous interactions; the coefficient of the *Patent Engineer: Prior Interactions with Evaluator* variable is positive and significant ($B = 0.52, p < 0.01$). However, this might have not been the case for patent inventors. As I argued in Appendix B, patent engineers were less likely to consult only patent inventors as they accumulated evaluation experience; while these results show that they preferred evaluators over inventors as the coefficient of the *Inventor: Consulted* dummy variable is negative and significant.

Finally, two interesting results are related to evaluators' two other dimensions of expertise besides depth. The coefficients of the evaluators' total experience and experience breadth are negative and significant ($B = -0.01, p < 0.01$, $B = -0.05, p < 0.01$). Regarding evaluators' total experience, I could speculate, on the one hand, that patent engineers may regard their contributions to evaluations more valuable as evaluators acquire experience and, supposedly, competence. On the other hand, evaluators may also have acquired an internal high level of reputation or status that might add to the social costs that patent engineers incur when they ask the contributions of other evaluators (Nebus, 2006). Regarding evaluators' expertise breadth, I could instead speculate that patent engineers perceive distributions of expertise characterized by greater breadth as being more diluted across knowledge domain and less focused on the focal technology. Indeed, these results are worth a deeper analysis that is beyond the scope of this study.

REFERENCES

Adner, R. and D.A. Levinthal
2001 Demand heterogeneity and technology evolution: implications for product and process innovation. Management Science, 47: 611-628.

Adner, R., and D.A. Levinthal
2004 What is not a real option: Considering boundaries for the application of real options to business strategy. Academy of Management Review, 29: 74-85.

Adner, Ron, Felipe A Csaszar, and Peter B Zemsky.
2014 'Positioning on a multiattribute landscape', Management Science, 60: 2794-815.

Adner, Ron, Laszlo Polos, Michael Ryall, and Olav Sorenson
2009 The case for formal theory', Academy of Management Review, 34: 201-08.

Argote, L., and E. Miron-Spektor
2011 "Organizational Learning: From Experience to Knowledge", Organization Science, 22: 1123-37

Ashby W.R.
1956 Introduction to Cybernetics. New York, NY: Wiley.

Axelrod, R.
2015 Structure of decision: The cognitive maps of political elites. Princeton university press.

Bae, J., G. C. Biddle, and C. W. Park
2022 "Managerial learning from analyst feedback to voluntary capex guidance, investment efficiency, and firm performance." Management Science 68: 583-607.

Bandura, A.
1977 Social learning theory. Prentice-Hall.

Baron, R. M., and D. A. Kenny
1986 "The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations." Journal of Personality and Social Psychology 51: 1173.

Barr, P. S., J. L. Stimpert, and A. S. Huff.
1992 "Cognitive change, strategic action, and organizational renewal." Strategic Management Journal 13: 15-36.

Barsalou, Lawrence W.
1991 'Deriving categories to achieve goals.' in, Psychology of learning and motivation (Elsevier).

Barwise, Patrick, and Sean Meehan
2004 Simply better: Winning and keeping customers by delivering what matters most (Harvard Business Review Press).
Benner, Mary J, and Mary Tripsas. 2012. 'The influence of prior industry affiliation on framing in nascent industries: The evolution of digital cameras', Strategic Management Journal, 33: 277-302.

Bartunek, J. M.
1984 "Changing interpretive schemes and organizational restructuring: The example of a religious order." Administrative Science Quarterly: 355-372.

Bettis, R. A., and S. Hu
2018 “Bounded rationality, heuristics, computational complexity, and artificial intelligence.” In Behavioural Strategy in Perspective. Emerald Publishing Limited.

Biais, B., D. Hilton, K. Mazurier, and S. Pouget
2005 Judgemental overconfidence, self-monitoring, and trading performance in an experimental financial market. The Review of Economic Studies, 72: 287-312.

Bingham, C. B., and K. M. Eisenhardt
2011 "Rational heuristics: the ‘simple rules’ that strategists learn from process experience." Strategic Management Journal 32: 1437-1464.

Bohns, V. K., F. J. Flynn
2013 Underestimating our influence over others at work. Research in Organizational Behaviour, 33: 97–112

Borgatti, S.P. and R. Cross
2003 A relational view of information seeking and learning in social networks. Management Science, 49: 432-445.

Boudreau, K.J., E.C. Guinan, K.R. Lakhani, and C. Riedl
2016 Looking across and looking beyond the knowledge frontier: Intellectual distance, novelty, and resource allocation in science. Management Science, 62: 2765-2783.

Bower, G. H., and H. R. Ernest
1981 Theories of Learning. Englewood Cliffs, NJ: Prentice-Hall

Bramson, A., P. Grim, D.J. Singer, W.J. Berger, G. Sack, S. Fisher, C. Flocken, and B. Holman
2017 Understanding polarization: Meanings, measures, and model evaluation. Philosophy of Science, 84: 115-159.

Bourgeois III, L. J.
1985 "Strategic goals, perceived uncertainty, and economic performance in volatile environments." Academy of Management Journal 28: 548-573.

Brehmer, B.
1980 "In one word: Not from experience." Acta Psychologica 45: 223-241.

Brodbeck, F.C., R. Kerschreiter, A. Mojzisch, and S. Schulz-Hardt,
2007 Group decision making under conditions of distributed knowledge: The information asymmetries model. Academy of Management Review, 32: 459-479.

Bunderson, J.S. and K.M. Sutcliffe

2002 Comparing alternative conceptualizations of functional diversity in management teams: Process and performance effects. Academy of Management Journal, 45: 875-893.

Brunswik, E.
1952 The Conceptual Framework of Psychology. Chicago, IL: University of Chicago Press.

Cangelosi, V. E., and W. R. Dill
1965 "Organizational learning: Observations toward a theory." Administrative Science Quarterly: 175-203.

Casella, G. and R.L. Berger
2010 Statistical inference. Cengage Learning.

Cattani, Gino, and Simone Ferriani
2008 'A core/periphery perspective on individual creative performance: Social networks and cinematic achievements in the Hollywood film industry', Organization Science, 19: 824-44.

Chi, M. T. H., P. J. Feltovich, and R. Glaser
1981 "Categorization and representation of physics problems by experts and novices." Cognitive Science 5: 121-152.

Choi, J, and D. Levinthal
2022 "Wisdom in the Wild: Generalization and Adaptive Dynamics." Organization Science.

Christensen, M. and T. Knudsen
2010 Design of decision-making organizations. Management Science, 56: 71-89.

Collins, A. G. E., and M. J. Frank
2013 "Cognitive control over learning: creating, clustering, and generalizing task-set structure." Psychological Review 120: 190.

Conti, Raffaele, Alfonso Gambardella, and Myriam Mariani
2014 'Learning to be Edison: Inventors, organisations, and breakthrough inventions', Organization Science, 25: 833-49.

Cohen, W.M. and D.A. Levinthal
1990 Absorptive capacity: A new perspective on learning and innovation. Administrative Science Quarterly, 128-152.

Cooper, R.G.,
1990 Stage-gate systems: a new tool for managing new products. Business horizons, 33: 44-54.

Criscuolo, P., L. Dahlander, T. Grohsjean, and A. Salter,
2017 Evaluating novelty: The role of panels in the selection of R&D projects. Academy of Management Journal, 60: 433-460.

Criscuolo, P., L. Dahlander, T. Grohsjean, and A. Salter

2021 The sequence effect in panel decisions: Evidence from the evaluation of research and development projects. Organization Science, 32: 987-1008.

Csaszar, F. A.
2018 “What makes a decision strategic? Strategic representations.” Strategy Science. 3: 606-19.

Csaszar, F. A., and D. A. Levinthal
2016 "Mental representation and the discovery of new strategies." Strategic Management Journal 37: 2031-2049.

Csaszar, F.A., and J.P. Eggers
2013 Organizational decision making: An information aggregation view. Management Science, 59: 2257-2277.

Csaszar, F. A., and J. Ostler
2020 "A contingency theory of representational complexity in organizations." Organization Science 31: 1198-1219.

Cyert, R.M., and J.G. March
1963 A Behavioral Theory of the Firm. Englewood Cliffs, NJ: Prentice-Hall.

Dane, E
2010 "Reconsidering the trade-off between expertise and flexibility: A cognitive entrenchment perspective." Academy of Management Review 35: 579-603.

Denrell, J
2008 “Superstitious Behaviour as a Byproduct of Intelligent Adaptation.” in G. P. Hodgkinson, and W. H. Starbuck (eds.), The Oxford Handbook of Organizational Decision Making. Oxford Academic.

Denrell, J., C. Fang, and D. A. Levinthal
2004 "From T-mazes to labyrinths: Learning from model-based feedback." Management Science 50: 1366-1378.

Deutsch, M. and H.B. Gerard
1955 A study of normative and informational social influences upon individual judgment. The Journal of Abnormal and Social Psychology, 51: 629.

Dutton, J. E., and S. E. Jackson
1987 "Categorizing strategic issues: Links to organizational action." Academy of Management Review 12: 76-90.

Eagly, A.H. and S. Chaiken
1984 Cognitive theories of persuasion. In Advances in experimental social psychology. 17: 267-359. Academic Press.

Ecken, P., and R. Pibernik
2016 Hit or miss: What leads experts to take advice for long-term judgments? Management Science, 62: 2002–2021

Eggers, J. P.
2012 Falling Flat: Failed Investments and Technological Evolution (April 10, 2012). Administrative Science Quarterly. 57: 47-80

Edwards, J. R., and L. S. Lambert
2007 "Methods for integrating moderation and mediation: a general analytical framework using moderated path analysis." Psychological Methods 12: 1.

Eggers, J. P., and S. Kaplan
2009 "Cognition and renewal: Comparing CEO and organizational effects on incumbent adaptation to technical change." Organization Science 20: 461-477.
2013 "Cognition and capabilities: A multi-level perspective." Academy of Management Annals 7: 295-340.

Einhorn, H.J., R.M. Hogarth, and E. Klempner
1977 Quality of group judgment. Psychological Bulletin, 84: 158.

Einhorn, H. J., and R. M. Hogarth
1986 "Decision making under ambiguity." Journal of Business: 225-250.

Ericsson, K. A., and J. Smith
1991 Toward a general theory of expertise: Prospects and limits. Cambridge University Press..

Felin, Teppo, and Todd R Zenger
2017 'The theory-based view: Economic actors as theorists', Strategy Science, 2: 258-71.

Feltovich, P. J., M. J. Prietula, and K. A. Ericsson
2018 "Studies of expertise from psychological perspectives: Historical foundations and recurrent themes." In K. A. Ericsson, R. R. Hoffman, A. Kozbelt, and A. M. Williams (eds.), The Cambridge Handbook of Expertise and Expert Performance. Cambridge University Press, 59–83.

Finton, D. J.
2005 "When do differences matter? On-line feature extraction through cognitive economy." Cognitive Systems Research 6: 263-281.

Fiol, C. M., and M. A. Lyles
1985 "Organizational learning." Academy of Management Review 10: 803-813.

Fiol, C. M., and E. J. O'Connor
2003 "Waking up! Mindfulness in the face of bandwagons." Academy of Management Review 28: 54–70.

Fiske, S. T., and S. E. Taylor
1984 Social cognition. Reading, MA: Addison-Wesley.

Fleming, L. and O. Sorenson
2004 Science as a map in technological search. Strategic Management Journal, 25: 909-928.

Fox, C.R. and A. Tversky

1995 Ambiguity aversion and comparative ignorance. The quarterly Journal of Economics, 110: 585-603.

Freeman, C., and L. Soete
1997 The economics of industrial innovation. Psychology Press.

French, K.R. and J. Poterba
1991 Investor diversification and international equity markets. The American Economic Review, 81: 222.

Fuchs, C., F.J. Sting, M. Schlickel, and O. Alexy
2019 The ideator's bias: How identity-induced self-efficacy drives overestimation in employee-driven process innovation. Academy of Management Journal, 62: 1498-1522.

Gardenfors, Peter
2004 Conceptual spaces: The geometry of thought (MIT press).

Gary, M. S., and R. E. Wood
2011 "Mental models, decision rules, and performance heterogeneity." Strategic Management Journal 32: 569-594.

Gavetti, G.
2005 "Cognition and hierarchy: Rethinking the microfoundations of capabilities' development." Organization Science 16: 599-617.
2011 'PERSPECTIVE—Toward a Behavioral Theory of Strategy', Organization Science, 23: 267-85.

Gavetti, G., and D. A. Levinthal,
2000 Looking forward and looking backward: Cognitive and experiential search. Administrative science quarterly, 45:113-137.

Gavetti, G., D. A. Levinthal, and J. W. Rivkin
2005 "Strategy making in novel and complex worlds: The power of analogy." Strategic Management Journal 26: 691-712.

Gavetti, G. and M. Warglien
2015 A model of collective interpretation. Organization Science, 26: 1263-1283.

Gershman, S. J., and Y. Niv
2013 "Perceptual estimation obeys Occam's razor." Frontiers in Psychology 4: 623.

Gigerenzer, G., and D. G. Goldstein
1996 "Reasoning the fast and frugal way: models of bounded rationality." Psychological Review 103: 650.

Glynn, M. A., T. K. Lant, and F. J. Milliken
1994 "Mapping learning processes in organizations: A multi-level framework for linking learning and organizing." In C. Stubbart, J. R. Meindl, J. F.A. Porac, (eds.) Advances in Managerial Cognition and Organizational Information Processing, Greenwich, CT: JAI Press, 43-83

Gopnik, A., and L. Schulz, and L. E. Schulz
2007 Causal learning: Psychology, philosophy, and computation. Oxford University Press.

Glaser, M., T. Langer, and M. Weber
2007 On the trend recognition and forecasting ability of professional traders. Decision Analysis, 4: 176-193.

Greene, W. H.
2003 Econometric analysis. Pearson Education India.

Greve, H.
2003 Organizational learning from performance feedback: A behavioral perspective on innovation and change. Cambridge University Press.

Gulati, R.
1995 Does familiarity breed trust? The implications of repeated ties for contractual choice in alliances. Academy of Management Journal, 38: 85-112.

Guler, I.
2007 "Throwing good money after bad? Political and institutional influences on sequential decision making in the venture capital industry." Administrative Science Quarterly 52: 248-285.

Haas, M.R., P. Criscuolo, and G. George
2015 Which problems to solve? Online knowledge sharing and attention allocation in organizations. Academy of Management Journal, 58: 680-711.

Hackman, J.R.,
1990 Groups that work and those that don't. Jossey-Bass.

Hall, C.C., L. Ariss, and A. Todorov
2007 The illusion of knowledge: When more information reduces accuracy and increases confidence. Organizational Behavior and Human Decision Processes, 103: 277-290.

Hallen, B. L
2008 The causes and consequences of the initial network positions of new organizations: From whom do entrepreneurs receive investments? Administrative Science Quarterly. 53: 685-718.

Hannigan, T. R., R. F.J. Haans, K. Vakili, H. Tchalian, V. L. Glaser, M. S. Wang, S. Kaplan, and P. D. Jennings
2019 "Topic modeling in management research: Rendering new theory from textual data." Academy of Management Annals 13: 586-632.

Harris, M. J., and R. Rosenthal
1985 "Mediation of interpersonal expectancy effects: 31 meta-analyses." Psychological Bulletin 97: 363.

Haunschild, P. R., and B. N. Sullivan

2002 "Learning from complexity: Effects of prior accidents and incidents on airlines' learning." Administrative Science Quarterly 47: 609-643.

Hayes, A. F.
2013 Introduction to mediation, moderation, and conditional process analysis: A regression-based approach. Guilford Press.

Heckman, J.J.
1979 Sample selection bias as a specification error. Econometrica: Journal of the Econometric Society. 153-161.

Heit, E.
1994 "Models of the effects of prior knowledge on category learning." Journal of Experimental Psychology: Learning, Memory, and Cognition 20: 1264.

Higham, K, G. De Rassenfosse, and A. B. Jaffe
2021 "Patent quality: Towards a systematic framework for analysis and measurement." Research Policy 50: 104215.

Hogg, M. A., and B. Mullin
1999 "Joining groups to reduce uncertainty: Subjective uncertainty reduction and group identification." In D. Abrams and M. A. Hogg (eds.), Social identity and social cognition. Oxford, UK: Blackwell, 249-279.

Hollenbeck, J. R., D. R. Ilgen, D. J. Sego, J. Hedlund, D. A. Major, and J. Phillips
1995 Multilevel theory of team decision making: Decision performance in teams incorporating distributed expertise. Journal of applied psychology, 80: 292.

Huberman, G
2001 Familiarity breeds investment. The Review of Financial Studies, 14: 659-680.

Huff, A. S., and M. Jenkins
2002 Mapping strategic knowledge. Sage.

Johnson-Laird, P. N.
1983 Mental models: Towards a cognitive science of language, inference, and consciousness. Harvard University Press.

Kaplan, S,, and M. Tripsas
2008 "Thinking about technology: Applying a cognitive lens to technical change." Research Policy: 790-805.

Kahneman, D. and A. Tversky
1979 Prospect theory: An analysis of decision under risk. In Handbook of the fundamentals of financial decision making: Part I. 99-127.

Kapoor, R. and N.R. Furr
2015 Complementarities and competition: Unpacking the drivers of entrants' technology choices in the solar photovoltaic industry. Strategic Management Journal, 36: 416-436.

Katila, Riitta, and Gautam Ahuja
2002 'Something Old, Something New: A Longitudinal Study of Search Behavior and New Product Introduction', The Academy of Management Journal, 45: 1183-94.

Khanna, R., I. Guler, and A. Nerkar
2018 "Entangled decisions: Knowledge interdependencies and terminations of patented inventions in the pharmaceutical industry." Strategic Management Journal 39: 2439-2465.

King, G., M. Tomz, and J. Wittenberg
2000 Making the most of statistical analyses: Improving interpretation and presentation. American Journal of Political Science. 44: 347–361

King, G., and L. Zeng
2001 Logistic regression in rare events data. Political Analysis. 9: 137-163.

Knight, F
1921 Risk, Uncertainty and Profit. Hart, Schaffner, and Marx, New York.

Kiesler, S., and L. Sproull
1982 "Managerial response to changing environments: Perspectives on problem sensing from social cognition." Administrative Science Quarterly: 548-570.

Kluger, A. N., and A. DeNisi
1996 "The effects of feedback interventions on performance: a historical review, a meta-analysis, and a preliminary feedback intervention theory." Psychological Bulletin 119: 254.

Knudsen, T. and D.A. Levinthal
2007 Two faces of search: Alternative generation and alternative evaluation. Organization Science, 18: 39-54.

Konlechner, S., and V. Ambrosini
2019 "Issues and trends in causal ambiguity research: a review and assessment." Journal of Management 45: 2352-2386.

Krackhardt, D.
1988 Predicting with networks: Nonparametric multiple regression analysis of dyadic data. Social Networks. 10: 359-381.

Kruschke, J. K.
1992 "ALCOVE: An exemplar-based connectionist model of category learning." Psychological Review 1: 107-138.

Kruger J., and D. Dunning
1999 Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. J. Personality Soc. Psychol. 77(6):1121–1134.

Langer, E. J.
1989 Mindfulness. Cambridge, MA: Perseus Books.

Lamberts K., and D. Shanks
1997 Knowledge Concepts and Categories. Psychology Press.

Laughlin, P.R.
2011 Group problem solving. Princeton University Press.

Leahey, Erin
2007 'Not by productivity alone: How visibility and specialisation contribute to academic earnings', American Sociological Review, 72: 533-61.

Lee, S. and F.A. Csaszar
2020 Cognitive and structural antecedents of innovation: A large-sample study. Strategy Science, 5: 71-97.

Lee, C.J., C.R Sugimoto., G. Zhang, and B. Cronin
2013 Bias in peer review. Journal of the American Society for Information Science and Technology, 64: 2-17.

Levinthal, D. A.
1997 "Adaptation on rugged landscapes." Management Science 43: 934-950.
2011 "A behavioral approach to strategy—what's the alternative?." Strategic Management Journal 32: 1517-1523.

Levinthal, D. A., and C. Rerup
2006 "Crossing an apparent chasm: Bridging mindful and less-mindful perspectives on organizational learning." Organization Science 17: 502-513.
2021 "The plural of goal: Learning in a world of ambiguity." Organization Science, 32: 527-43

Levitt, B., and J. G. March
1988 "Organizational Learning", Annual Review of Sociology, 14: 319-38

Li, D.
2017 Expertise versus Bias in Evaluation: Evidence from the NIH. American Economic Journal: Applied Economics, 9: 60-92.

Li, H., S. Rosen, and W. Suen,
2001 Conflicts and common interests in committees. American Economic Review, 91: 1478-1497.

Lurigio, A. J., and J. S. Carroll
1985 "Probation officers' schemata of offenders: Content, development, and impact on treatment decisions." Journal of Personality and Social Psychology 48: 1112.

MacGeorge, E.L., B. Feng, and L.M. Guntzviller
2016 Advice: Expanding the communication paradigm. Annals of the International Communication Association, 40: 213-243.

Mannucci, P.V. and K. Yong,

2018 The differential impact of knowledge depth and knowledge breadth on creativity over individual careers. Academy of Management Journal, 61: 1741-1763.

Martins, L.L. and W. Sohn,
2022 How does diversity affect team cognitive processes? Understanding the cognitive pathways underlying the diversity dividend in teams. Academy of Management Annals, 16: 134-178.

March, J. G.
2010 The ambiguities of experience. Cornell University Press.

March, J. G., and H. A. Simon
1958 Organizations. New York, NY: Wiley.

March, J. G., L. S. Sproull, and M. Tamuz
1991 "Learning from samples of one or fewer." Organization Science 2: 1–13.

Martignoni, D., A. Menon, and N. Siggelkow
2016 "Consequences of misspecified mental models: Contrasting effects and the role of cognitive fit." Strategic Management Journal 37: 2545-2568.

McNamara, G. M., R. A. Luce, and G. H. Tompson
2002 "Examining the effect of complexity in strategic group knowledge structures on firm performance." Strategic Management Journal 23: 153-170.

Michel, A. A.
2007 "A distributed cognition perspective on newcomers' change processes: The management of cognitive uncertainty in two investment banks." Administrative Science Quarterly 52: 507-557.

Miller, C.C., S. Chiu, C.L. Wesley II, D. Vera, D.R. Avery,
2022 Cognitive Diversity at The Strategic Apex: Assessing Evidence on the Value of Different Perspectives and Ideas among Senior Leaders. Academy of Management Annals, 16: 806-852.

Milliken, F.J.
1987 Three types of perceived uncertainty about the environment: State, effect, and response uncertainty. Academy of Management Review, 12: 133-143.

Miner, A., and R. Haunschild
1995 "Population level learning." In L. L. Cummings and B. M. Staw (eds.), Research in Organizational Behavior, 17: 115-166.

Morris, M. W., and P. C. Moore
2000 "The lessons we (don't) learn: Counterfactual thinking and organizational accountability after a close call." Administrative Science Quarterly 45: 737-765.

Mosakowski, E.
1997 "Strategy making under causal ambiguity: Conceptual issues and empirical evidence." Organization Science 8: 414-442.

Murphy, G. L.
2004 The big book of concepts. MIT press.

Murphy, G. L., and H. H. Brownell
1985 "Category differentiation in object recognition: typicality constraints on the basic category advantage." Journal of Experimental Psychology: Learning, Memory, and Cognition 11: 70.

Murphy, G. L., and D. L. Medin
1985 "The role of theories in conceptual coherence." Psychological Review 92: 289.

Nebus, J.,
2006 Building collegial information networks: A theory of advice network generation. Academy of Management Review, 31: 615-637.

Nelson, R. R.
2008 "Bounded rationality, cognitive maps, and trial and error learning." Journal of Economic Behavior & Organization 67: 78-89.

Nelson, R., and S. Winter
1982 An Evolutionary Theory of the Firm. Cambridge, MA: Harvard University Press.

Nisbett, R.E., and L. Ross
1980 Human Inference : Strategies and Shortcomings of Social Judgment. Englewood Cliffs, NJ: Prentice-Hall.

Nolan, J. M., P. W. Schultz, R. B. Cialdini, N. J. Goldstein, and V. Griskevicius
2008 Normative social influence is underdetected. Personality & Social Psychology Bulletin, 34: 913–923.

Oinas-Kukkonen, H.
2008 Network analysis and crowds of people as sources of new organisational knowledge. In: A. Koohang et al. (Eds): Knowledge Management: Theoretical Foundation. Informing Science Press, Santa Rosa, CA, US. 173–189.

Owen, G. and B. Grofman
1986 Information pooling and group decision making. Greenwich, CT: JAI.

Packard, M.D., B.B. Clark, and P.G. Klein
2017 Uncertainty types and transitions in the entrepreneurial process. Organization Science, 28: 840-856.

Paich, M., and J. D. Sterman
1993 "Boom, bust, and failures to learn in experimental markets." Management Science 39: 1439-1458.

Porac, J. F., H. Thomas, F. Wilson, D. Paton, and A. Kanfer
1995 "Rivalry and the industry model of Scottish knitwear producers." Administrative Science Quarterly: 203-227.

Preacher, K. J., D. D. Rucker, and A. F. Hayes

2007 "Addressing moderated mediation hypotheses: Theory, methods, and prescriptions." Multivariate Behavioral Research 42: 185-227.

Rader, C.A., R.P. Larrick, and J.B. Soll
2017 Advice as a form of social influence: Informational motives and the consequences for accuracy. Social and Personality Psychology Compass, 11.

Radulescu, A., Y. Niv, and I. Ballard
2019 "Holistic reinforcement learning: the role of structure and attention." Trends in Cognitive Sciences 23: 278-292.

Radulescu, A., Y. S. Shin, and Y. Niv
2021 "Human representation learning." Annual Review of Neuroscience 44: 253-273.

Rahmandad, H.
2008 "Effect of delays on complexity of organizational learning." Management Science 54: 1297-1312.

Reagans, R., L. Argote, and D. Brooks
2005 "Individual experience and experience working together: Predicting learning rates from knowing who knows what and knowing how to work together." Management Science 51: 869-881.

Rehder, B.
2003 "Categorization as causal reasoning." Cognitive Science 27: 709-748.

Ren, Y. and L. Argote
2011 Transactive memory systems 1985–2010: An integrative framework of key dimensions, antecedents, and consequences. Academy of Management Annals, 5: 189-229.

Roberts, J.H. and J.M. Lattin
1991 Development and testing of a model of consideration set composition. Journal of Marketing Research, 28: 429-440.

Rogers, E.M.
1995 Diffusion of innovations. Simon and Schuster.

Rosch, E., C. B. Mervis, W. D. Gray, D. M. Johnson, and P. Boyes-Braem
1976 "Basic objects in natural categories." Cognitive Psychology 8: 382-439.

Rosch, E., and B. B. Lloyd
1978 Cognition and categorization. Hillsdale, NJ: Erlbaum

Sah R.K., J.E. Stiglitz
1986 The architecture of economic systems: Hierarchies and polyarchies. American Economic Review 76: 716–727.

Serrano, C. J.
2010 "The dynamics of the transfer and renewal of patents." The RAND Journal of Economics 41: 686-708.

Shaver, J. M.
2005 "Testing for mediating variables in management research: Concerns, implications, and alternative strategies." Journal of Management 31: 330-353.

Sherif, M
1935 A study of some social factors in perception. Archives of Psychology, 27: 187

Simon, H. A.
1962 "The Architecture of Complexity." Proceedings of the American Philosophical Society 106: 467-482.
1990 "Invariants of human behavior." Annual Review of Psychology 41: 1-20.
1991 "Bounded rationality and organizational learning." Organization Science 2: 125-134.

Stasser, G., N. L. Kerr, and J. H. Davis
1989 Influence processes and consensus models in decision-making groups. In P. B. Paulus (Ed.), Psychology of group influence: 279 326. Hillsdale, NJ: Lawrence Erlbaum Associates

Stasser, G., and Z. Birchmeier
2003 Group creativity and collective choice. In P. B. Paulus & B. Nijstad (Eds.), Group creativity: 132-172. New York: Oxford University Press

Stasser, G., and W. Titus
1985 Pooling of unshared information in group decision making: Biased information sampling during discussion. Journal of Personality and Social Psychology, 48: 1467.

Sterman, J. D.
1989 "Modeling managerial behavior: Misperceptions of feedback in a dynamic decision-making experiment." Management Science 35: 321-339.

Sull, D.N., and K. M. Eisenhardt
2015 Simple Rules: How to Thrive in a Complex World. Boston, MA: Houghton Mifflin Harcourt.

Sundstrom, E., K.P. De Meuse, and . Futrell,
1990 Work teams: Applications and effectiveness. American psychologist, 45: 120.

Surowiecki, J.
2004 The wisdom of crowds: Why the many are smarter than the few and how collective wisdom shapes business. Economies, Societies and Nations, 296.

Szulanski, G., R. Cappetta, and R.J. Jensen,
2004 When and how trustworthiness matters: Knowledge transfer and the moderating effect of causal ambiguity. Organization Science, 15: 600-613.

Tanaka, J. W., and M. Taylor
1991 "Object categories and expertise: Is the basic level in the eye of the beholder?." Cognitive Psychology 23: 457-482.

Taylor, A. and H.R. Greve,
2006 Superman or the fantastic four? Knowledge combination and experience in innovative teams. Academy of Management Journal, 49: 723-740.

Teece, D., M. Peteraf, and S. Leih
2016 Dynamic capabilities and organizational agility: Risk, uncertainty, and strategy in the innovation economy. California Management Review, 58: 13-35.

Tenenbaum, J. B., C. Kemp, T. L. Griffiths, and N. D. Goodman
2011 "How to grow a mind: Statistics, structure, and abstraction." Science 331: 1279-1285.

Teodoridis, Florenta, Michaël Bikard, and Keyvan Vakili
2019 'Creativity at the knowledge frontier: The impact of specialisation in fast-and slow-paced domains', Administrative Science Quarterly, 64: 894-927.

Thagard, P.
2005 "Representation and computation." In P. Thagard (ed.) Mind: Introduction to cognitive science: 3-22.

Thompson, J.D.
1967 Organizations in action: Social science bases of administrative theory. Transaction publishers.

Tindale, R.S., T. Kameda, and V.B. Hinsz
2003 Group decision making. Sage handbook of social psychology. 381-403.

Toh, P.K. and T. Kim
2013 Why put all your eggs in one basket? A competition-based view of how technological uncertainty affects a firm's technological specialization. Organization Science, 24: 1214-1236.

Tomz, M
2003 Relogit: Rare events logistic regression. Journal of Statistical Software. 8: 137–163

Tripsas, M., and G. Gavetti
2000 "Capabilities, cognition, and inertia: evidence from digital imaging." Strategic Management Journal, 21: 1147-1161

Tsoukas, H.
2009 "A dialogical approach to the creation of new knowledge in organizations." Organization Science 20: 941-957.

Van de Ven, A.H. and R. Garud
1993 Innovation and industry development: The case of cochlear implants. Research on Technological innovation, Management and Policy, 5: 1-46.

Wang, B., and C.-H. Hsieh
2015 "Measuring the value of patents with fuzzy multiple criteria decision making: insight into the practices of the Industrial Technology Research Institute." Technological Forecasting and Social Change: 263-275.

Walker, G.
1985 "Network position and cognition in a computer software firm." Administrative Science Quarterly: 103-130.

Walsh, J. P.
1995 "Managerial and organizational cognition: Notes from a trip down memory lane." Organization Science 6: 280-321.

Wang, B., and C.-H. Hsieh.
2015 "Measuring the value of patents with fuzzy multiple criteria decision making: insight into the practices of the Industrial Technology Research Institute." Technological Forecasting and Social Change: 263-275.

Weick, K. E.
1979 "Cognitive Processes in Organizations," In B. Staw (ed.) Research in Organizational Behavior, Greenwich, CT: JAI Press, 1: 41-74

Weick, K. E., and K. H. Roberts
1993 “Collective mind in organization: Heedful interrelating on flight decks.” Administrative Science Quarterly 38: 357–381.

Weick, K. E., K. M. Sutcliffe, and D. Obstfeld
1999 “Organizing for high reliability: Processes of collective mindfulness.” Research in Organizational Behavior 21: 81–123.

Wood, W
2000 Attitude change: Persuasion and social influence. Annual Review of Psychology, 51: 539–570.

Wooldridge, J.M.
2010 Econometric analysis of cross section and panel data. MIT press.

Zelner, B. A.
2009 Using simulation to interpret results from logit, probit, and other nonlinear models. Strategic Management Journal. 30: 1335–1348.

Zollo, M.
2009 "Superstitious learning with rare strategic decisions: Theory and evidence from corporate acquisitions." Organization Science 20: 894-908.

Zollo, M., and S. G. Winter
2002 "Deliberate learning and the evolution of dynamic capabilities." Organization Science 13: 339-351.

Zott, C. and R. Amit
2008 The fit between product market strategy and business model: Implications for firm performance. Strategic Management Journal, 29: 1-26.