



IE UNIVERSIDAD

TESIS DOCTORAL

**MEJORA DE LA PRECISIÓN DE PREDICCIONES
COLECTIVAS VIA ORGANIZACIÓN CONSCIENTE**

EFRAIN ROSEMBERG MONTES

ASESOR TESIS DOCTORAL: MATTHIAS SEIFERT

SEGOVIA, 2024



IE UNIVERSIDAD

DOCTORAL DISSERTATION

IMPROVING GROUP FORECAST ACCURACY VIA
MINDFUL ORGANIZING

EFRAIN ROSEMBERG MONTES

DOCTORAL THESIS ADVISOR: MATTHIAS SEIFERT

SEGOVIA, 2024

RESUMEN (TRADUCCIÓN AL ESPAÑOL)

En las últimas cuatro décadas, los investigadores han reconocido que comprender los aspectos organizativos de la disciplina de *forecasting* podría arrojar luz sobre los desafíos frecuentemente citados en el campo: mejorar la precisión de las predicciones, implementar técnicas de previsión y lograr una alineación entre las funciones de las empresas. A pesar de esta promesa, la dinámica organizativa en *forecasting* sigue estando inexplorada en gran medida, posiblemente debido a un énfasis en el paradigma funcionalista que postula que los *forecasters* ayudan a la toma de decisiones al predecir una realidad concreta "allí afuera" que existe independientemente de interacciones y las creencias de los participantes en el proceso de *forecast*. Dentro de este contexto, esta tesis propone un paso paradigmático hacia una "zona de transición" funcionalista-interpretativa utilizando el marco de organización consciente como puente (Weick, 1995; Weick et al., 1999). Este marco ofrece una plantilla matizada de interacciones de la vida real centrada en la calidad de la atención organizativa, el trabajo interpretativo colectivo y una sensibilidad particular para analizar errores y quasi accidentes. Manteniendo el objetivo de precisión, este trabajo puede contribuir al campo de la predicción desde tres perspectivas: a) proporcionando información sobre los microprocesos de comunicación en equipos que podrían reducir los sesgos en el juicio, b) enfatizando el papel de la divulgación de errores y la deliberación en el rendimiento de equipos, y c) fundamentando teóricamente la subjetividad inherente del proceso de *forecasting*, donde los participantes interactúan activamente para predecir y decidir. Informamos los hallazgos empíricos de un estudio experimental que involucra a 40 tríadas jerárquicas de *forecast*, un líder y dos miembros del equipo, para probar el papel de la organización consciente en *forecasting*. En comparación con el grupo de control, encontramos mejoras sistemáticas en la precisión de *forecast* de valores puntuales cuando los grupos han recibido intervenciones para fomentar interacciones alineadas con el marco de organización consciente. Sin embargo, la organización consciente no mejora la precisión de *forecasts* de intervalo (MSIS). Analizando las interacciones de los equipos a través de la funcionalidad de chat arroja luz sobre una comunicación más amplia y matizada en los grupos de intervención, fomentando la participación y las opiniones opuestas entre los miembros del equipo. Se discuten las implicaciones para la teoría y la práctica del *forecasting* en entornos organizacionales.

ABSTRACT

In the last four decades, researchers have acknowledged that understanding the organizational aspects of forecasting could shed light on often-cited challenges in the field: improving prediction accuracy, implementing forecasting techniques, and achieving alignment between firms' functions. Despite this pledge, organizational dynamics in forecasting remain largely uncharted, possibly due to an over-emphasis on the functionalistic paradigm that posits forecasters aid decision-making by predicting a concrete reality "out there" that exists independently of the participants' interactions and beliefs. Within this context, this thesis proposes a paradigmatic step towards a functionalist-interpretive "transition zone" using the mindful organizing framework as a bridge (Weick, 1995; Weick et al., 1999). This framework conveys a nuanced template of real-life interactions focused on collective interpretive work and a particular sensitivity to analyze errors and near misses. While retaining the objective goal of accuracy, this work can contribute to the forecasting field from three angles: a) providing insights into team communication micro-processes that could debias judgment, b) emphasizing the role of error disclosure and deliberation in team performance, and c) substantiating theoretically the inherent subjectivity of the forecast process where participants interact actively to predict and decide. We report the empirical findings of an experimental study involving 40 hierarchical forecasting triads—one leader and two team members—to test the role of mindful organizing in judgmental forecasting. Compared to the control group, we find systematic improvements in point forecast accuracy when groups have received interventions to foster interactions aligned with the mindful organizing framework. However, mindful organizing in forecasting teams does not improve interval accuracy scores (MSIS). Analyzing the teams' interactions via chat functionality sheds light on more extended and nuanced communication in the intervention groups, prompting participation and opposing views amongst team members. Implications for the theory and practice of forecasting in organizational environments are discussed.

ACKNOWLEDGEMENTS

I extend my deepest gratitude to the following individuals whose support and inspiration have been instrumental in my academic journey:

My Father: His belief in education and his pursuit of knowledge ignited my passion for doctoral studies. His guidance and encouragement have been invaluable.

My Mother: A beacon of resilience and commitment, my mother's dedication to her aspiration for her family has been a constant source of motivation. Her example taught me the true meaning of perseverance.

My Wife: Amid the challenges of balancing studies and family life, my wife stood by me with relentless support. Her encouragement, patience, and belief in my abilities fueled my determination to complete this dissertation.

Their love, sacrifices, and unwavering faith have shaped my journey; I am forever grateful.

LIST OF CONTENTS

ABSTRACT (SPA/ENG)	I
ACKNOWLEDGEMENTS	III
LIST OF CONTENTS	IV
List of Figures	VI
List of Tables	VII
I. DISSERTATION INTRODUCTION (SPA/ENG)	1
II. THEORETICAL CHAPTER	7
1. INTRODUCTION	7
2. THE EFFECTS OF FUNCTIONALISM IN FORECASTING RESEARCH ...	11
3. ORGANIZATIONAL NUANCES IN FORECASTING	15
4. MINDFUL ORGANIZING FRAMEWORK	23
5. CONCLUSION	31
III. EMPIRICAL CHAPTER	33
1. INTRODUCTION	33
2. THEORETICAL BACKGROUND	37
2.1. Group forecast extraction and task feedback.....	37
2.2. Team characteristics	39
3. PROPOSED RESEARCH MODEL AND HYPOTHESES	41
3.1. Debiasing group judgment with mindful organizing	41
3.2. Elicitation of mindful organizing: team engagement design	43
4. METHODS	47
4.1. Task Design & Data Sample	47
4.2. Experimental Protocol	49

4.2.1. Time series data generation	52
4.3. Operationalization of Key Constructs	53
5. DATA ANALYSIS AND RESULTS	56
5.1. Manipulation check	56
5.2. Measuring team accuracy using leaders' judgment	57
5.3. Measuring team accuracy by aggregating participants' forecasts	59
5.4. Robustness checks	61
5.5. Team chat analysis	64
6. DISCUSSION AND CONCLUSIONS.....	66
IV. DISCUSSION AND CONCLUSIONS (SPA/ENG)	70
V. APPENDIX.....	76
1. Instructions.....	76
2. Intervention videos.....	77
3. Forecasting game interfaces.....	81
4. Q-Q plots for normality tests>.....	82
VI. REFERENCES	83

LIST OF FIGURES

Figure 1: Theoretical model.....	27
Figure 2: Forecasting exercise screen layouts.....	51
Figure 3: Time series including team forecasts and real values (last five values).....	52
Figure 4: SOS items and correspondence with Mindful Organizing from Vogus & Sutcliffe (2007)	55
Figure 5: SOS in control and intervention groups.....	56
Figure 6: Pairwise mean difference and dispersion for accuracy metrics	63

LIST OF TABLES

Table 1: Mindful organizing counteracts group judgment propensities ...	43
Table 2: Sample age distribution	48
Table 3: Measures of accuracy	53
Table 4: Experimental conditions accuracy	58
Table 5: Regression results, p values<0.05	59
Table 6: T-tests for different aggregation metrics	60
Table 7: Regression analyses for MAPE avg group, MSIS6 and MSIS4	60
Table 8: Multivariate tests, experimental condition	62
Table 9: Multivariate tests, dependent variables	63

I. INTRODUCCION DE LA TESIS (traducción al español)

Aunque es fundamental para la toma de decisiones y la planificación, el *forecasting* enfrenta desafíos persistentes relacionados con la precisión, a menudo atribuidos a sesgos de juicio colectivo y dinámicas organizativas. A pesar de reconocer la importancia del contexto sociotécnico, la investigación existente ha adoptado en gran medida perspectivas funcionalistas, tratando a los *forecasters* como predictores objetivos aislados del contexto organizativo. No sorprendentemente, este enfoque pasa por alto la intrincada interacción de la jerarquía grupal y la coordinación del equipo, la subjetividad, la cognición colectiva y la gestión de errores dentro de los equipos de pronóstico. La premisa general de este estudio es infundir teoría relevante de organización y toma de decisiones en la predicción para afinar nuestra comprensión de estos fenómenos y mejorar nuestra capacidad para hacer recomendaciones

El estudio se ha dividido en dos capítulos: un segmento teórico más exploratorio y un segmento empírico que comprende un estudio experimental. En su capítulo teórico, este estudio explora las revisiones de la literatura de los últimos 40 años para conectar los llamados dispersos a estudios que enfatizan los aspectos organizativos del *forecasting*. Una conclusión de este análisis es que las limitaciones percibidas en la comprensión de la predicción desde un punto de vista organizativo podrían estar relacionadas con el enraizamiento paradigmático, un énfasis inadvertido en herramientas funcionalistas que se centran en lograr un mejor modelado, pero pasan por alto los aspectos humanos de las organizaciones como fenómenos marginales, lo que dificulta la incorporación de teoría relevante de toma de decisiones y organización en el campo. En consecuencia, este estudio busca abordar la brecha en la literatura profundizando en algunos aspectos organizativos superficialmente cubiertos por el enfoque dominante, como la extracción de pronósticos de grupos y la incorporación de retroalimentación de tareas. Además, aborda áreas pasadas por alto en la literatura de *forecasting*, como la deliberación del rendimiento, la facilitación del liderazgo y el aprendizaje de los errores. Finalmente, este

segmento teórico introduce el concepto de organización consciente como un puente entre los paradigmas funcionalista e interpretativo en la predicción. La organización consciente enfatiza la calidad de la atención organizativa, la construcción colectiva de sentido y la sensibilidad a los errores. La conclusión subyacente de esta parte es que al examinar e intervenir en cómo los equipos deliberan, revelan errores y aprenden colectivamente, las tareas de pronóstico pueden mejorarse.

En su capítulo empírico, el estudio presenta un experimento controlado aleatorio que involucra a equipos de pronóstico para investigar si un constructo como la organización consciente genera diferencias significativas en métricas de precisión de pronóstico colectivo, que van desde un pronóstico puntual del equipo (representado por la decisión final de los líderes) hasta la agregación de pronósticos del equipo y medidas de intervalo. Los participantes formaron tríadas con un líder y dos participantes, y se dedicaron a tareas de pronóstico utilizando una serie temporal. El grupo de intervención recibió capacitación en organización consciente, centrada en estrategias de comunicación y facilitación del equipo, mientras que los grupos de control vieron videos de distracción. La intervención de organización consciente mejoró significativamente los pronósticos del líder del equipo (medidos por MAPE leader) en un 24%. No se observaron efectos significativos en las medidas de intervalo. Adicionalmente, la comunicación del equipo se analizó utilizando la funcionalidad de chat en texto, mostrando patrones de comunicación más extensos y ricos en los grupos de intervención

Finalmente, se presentan posibles direcciones para futuras investigaciones. Una investigación adicional podría explorar factores mediadores o moderadores potenciales, como la cohesión del equipo, los patrones de comunicación o las características de los miembros del equipo que pueden interactuar con el rendimiento en tareas de pronóstico. Comprender estos mecanismos subyacentes podría proporcionar una comprensión más completa de cómo las intervenciones de atención plena afectan el rendimiento del equipo. Los estudios futuros también podrían examinar la sostenibilidad de la intervención a lo largo de un período prolongado, especialmente dado que no hay consenso en la

literatura sobre cuánto tiempo lleva que la organización consciente emerja y si puede mantenerse. Los estudios longitudinales o las evaluaciones de seguimiento podrían ayudar a determinar si las mejoras observadas en la precisión del equipo persisten con el tiempo o disminuyen después del período de intervención. Por último, explorar las condiciones límite de la efectividad de la intervención puede mejorar aún más la comprensión de su aplicabilidad.

El esquema presentado abajo contiene la estructura general de la tesis incluyendo la pregunta de investigación general y la estructura específica de los capítulos teóricos y empíricos.

<p>Pregunta de investigación principal: ¿Cómo podemos mejorar la precisión de los pronósticos colectivos al mejorar las interacciones del equipo?</p>	
<p>Capítulo teórico</p>	<p>Capítulo empírico</p>
<p><u>Pregunta de investigación:</u> ¿Cómo podemos infundir más teoría organizativa en el campo del <i>forecasting</i>?</p> <p><u>Método:</u> Revisión de la literatura centrada en a) revisiones de <i>forecasting</i> de los últimos 40 años, b) estudios de pronóstico que abordan elementos organizativos, c) literatura sobre aprendizaje organizativo y d) literatura sobre alta confiabilidad, sentido y organización consciente</p> <p><u>Hallazgos:</u></p> <ul style="list-style-type: none"> • Identificación de los efectos limitantes del paradigma existente (funcionalismo) en algunas corrientes de investigación de <i>forecasting</i> relacionadas con organizaciones y equipos, como la extracción de pronósticos de grupos y el aprendizaje a partir de la retroalimentación de pronósticos. • Reconocimiento de que los marcos potenciales deben introducir matices organizativos, como la naturaleza jerárquica de los equipos, la importancia del rol del líder, la deliberación del rendimiento a través de rituales posteriores y la relevancia de revelar errores. • Selección de la organización consciente como un marco teórico adaptativo potencial que se centra en la calidad de la atención organizativa, los microprocesos de comunicación y la deliberación. 	<p><u>Pregunta de investigación:</u> ¿Cómo podemos mejorar la precisión de los pronósticos colectivos utilizando los principios de la organización consciente?</p> <p><u>Método:</u> Estudio experimental que incluye tríadas jerárquicas asignadas al azar a condiciones de intervención y control.</p> <p><u>Hallazgos:</u></p> <ul style="list-style-type: none"> • La intervención de organización consciente mejoró significativamente los pronósticos del líder del equipo (medidos por MAPE leader) en un 24%. Otras métricas de precisión puntual también mostraron resultados significativos, pero no se observaron efectos significativos en las medidas de intervalo.

I. DISSERTATION INTRODUCTION

While critical for decision-making and planning, forecasting faces persistent challenges related to accuracy, often attributed to collective judgment biases and organizational dynamics. Despite acknowledging the significance of the socio-technical context, extant research has largely embraced functionalist perspectives, treating forecasters as objective predictors isolated from the organizational context. This approach has devoted less attention to the intricate interplay of team coordination and group hierarchy, subjectivity, collective cognition, and error management within forecasting teams. Hence, the overall premise of this study is to infuse relevant organization and decision-making theory into forecasting to refine our understanding of these phenomena and improve our ability to make prescriptions.

The study has been divided into two chapters: a more exploratory theoretical segment and an empirical segment comprising an experimental study. In its theoretical chapter, this study explores the literature reviews of the past 40 years to connect the intermittent calls for studies emphasizing organizational aspects of forecasting. One lead of this analysis is that the perceived limitations in understanding forecasting from an organizational point of view might be related to paradigmatic entrenchment, an inadvertent emphasis on functionalistic tools that default on striving for better modeling but disregard the human aspects of organizations as fringe phenomena, making it difficult to infuse relevant decision making and organizational theory in the field. Consequently, this study seeks to address the literature gap by revisiting some organizational aspects covered by the dominant approach, such as extracting forecasts from groups and incorporating task feedback. Also, it covers overlooked areas in the forecasting literature, such as performance deliberation, leader facilitation, and learning from mistakes. Finally, this theoretical segment introduces the concept of mindful organizing as a bridge between functionalist and interpretive paradigms in

forecasting. A mindful organizing framework emphasizes group interactions, the quality of organizational attention, collective sensemaking, and sensitivity to errors. The underlying conclusion of this part is that forecasting tasks can be enhanced by examining and intervening in how teams deliberate, disclose errors, and learn collectively.

In its empirical chapter, the study lays out a randomized controlled experiment involving forecasting teams to investigate whether the mindful organizing construct creates significant differences between intervention and control groups in metrics of collective forecast accuracy, including a team point forecast—represented by the leaders' decisions—and metrics of team forecast aggregations such as group mean and interval forecasts. Participants formed triads with one leader and two participants and engaged in forecasting tasks using a time series. The intervention group received mindful organizing training, focused on team communication and facilitation strategies, while the control groups watched decoy videos. The mindful organizing intervention significantly improved team leader forecasts (measured by MAPE leader) by 24%. Other point accuracy metrics also showed significant results, but no significant effects were observed on the interval measures. Additionally, team communication is analyzed using text-in-chat functionality, which shows longer and richer communication patterns in the intervention groups.

Finally, future research avenues are presented. Further investigation could explore potential mediating or moderating factors such as team cohesion, communication patterns, or team members' characteristics that may interact with forecasting task performance. Understanding these underlying mechanisms could provide a more comprehensive understanding of how mindfulness interventions impact team performance. Future studies could also examine the sustainability of the intervention over an extended period, especially since there is no consensus in the literature about how long it takes mindful organizing to emerge and whether it can be sustained. Longitudinal studies or follow-up assessments could help determine whether the observed improvements in team accuracy persist over time or diminish after the intervention period. Finally,

exploring the boundary conditions of the intervention's effectiveness can further enhance the understanding of its applicability.

The scheme presented below contains the general structure of the thesis, including the overall research question and the specific structure of the theoretical and empirical chapters.

<p>Main research question: how can we improve collective forecast accuracy by improving team interactions?</p>	
<p>Theoretical chapter</p>	<p>Empirical Chapter</p>
<p>Research question: How can we infuse more organizational theory into the forecasting field?</p> <p>Method: Literature review focusing on a) forecast reviews of the last 40 years, b) forecasting studies addressing organizational elements, c) literature on organizational learning, and d) literature on high reliability, sensemaking, and mindful organizing</p> <p>Findings:</p> <ul style="list-style-type: none"> • Identification of the limiting effects of the extant paradigm—functionalism—in some forecasting research streams situated in the adjacency of organizations and teams, such as extracting forecasts from groups and learning from forecast feedback. • Acknowledgment that potential frameworks must introduce organizational nuances such as the hierarchical nature of teams, the importance of the leader's role, performance deliberation through after-the-fact rituals, and the importance of disclosing mistakes. • Selection of Mindful organizing as a potential adaptive theoretical frame that focuses on the quality of organizational attention, communication micro-processes, and a focus on deliberation. 	<p>Research Question: How can we improve collective forecast accuracy using mindful organizing principles?</p> <p>Method: Experimental study including hierarchical triads randomly assigned to intervention and control conditions</p> <p>Findings:</p> <ul style="list-style-type: none"> • The mindful organizing intervention significantly improved team leader forecasts (measured by MAPE leader) by 24%. Other point accuracy metrics also showed significant results, but no significant effects were observed on the interval measures.

II. THEORETICAL CHAPTER

1. INTRODUCTION

In Rosemberg (2023), I highlight that some forecasting domains have experienced poor predictive accuracy beyond uncertainty's expected effects (Kahneman & Lovallo, 1993; Makridakis et al., 2010; Makridakis & Taleb, 2009). Outside the frequently reviewed challenges of demand forecast accuracy (Goodwin et al., 2018; Lawrence et al., 2006; Stewart, 2001), practitioners and academics have reported similar outcomes in other domains. For instance, a large-scale study on 5,400 information technology (IT) projects with budgets above \$15M concluded that 45% of projects exceeded their budgets and achieved 56% less value than forecasted (Bloch et al., 2012). Consistent with this, a 2015 report from the US Government Accountability Office (US GAO) offered support for those conclusions: *"Federal investments in information technology (IT) have often resulted in multimillion-dollar cost overruns and years-long schedule delays, with questionable mission-related achievements"* (GAO-15-675T, 2015, p.2). In a different industry, Nielsen's 2015 Breakthrough Innovation Report (Europe Edition) has outlined success rates of 0.2% (18 out of 8,650) in packaged goods launches. Remarkably, the lack of improvement over time persists, although evidence-based recommendations have been suggested repeatedly. Perhaps one of the most salient examples of inaccuracy recidivism comes from research on Mega Projects (budgets over \$1bn), where nine out of ten projects exceed budgets, with deviations of fifty percent being frequent (Flyvbjerg, 2014). The data reveals that these failure rates have remained constant for the last 70 years regardless of sectors and geographies.

In response to the mounting evidence, the field's literature reviews of the last four decades have identified the need for more research on the organizational aspects of forecasting to shed light on the difficulties of implementing forecasting techniques, the lack of forecast alignment between functions in firms, and the

absence of learning in some forecasting domains. In one of the first forecasting reviews referring to organizational aspects, Makridakis & Wheelwright (1977) highlighted the need for more attention to the environment's organizational characteristics. Subsequently, Hogarth & Makridakis (1981) suggested a reconceptualization of forecasting and planning (F&P) by adopting a multi-attribute analysis approach that evaluates F&P on organizational aspects (flexibility, motivation, control), while Armstrong, (1988) and Winklhofer et al. (1996) emphasized the need to improve our understanding of organizational challenges as a precondition to forecasting techniques implementation. Despite these calls, years later, Lawrence et al. (2006) have asserted that forecasts produced by groups have been largely neglected in their review of 25 years of judgmental forecasting, whereas Fildes (2006) has identified a gap in the contribution of forecasting journals to the understanding of organizational issues, which has resulted in a persistent gap between theory and forecasting practice.

According to Rosemberg (2023), from a research methods angle, Fildes et al. (2009) have made an explicit call for revisiting the methods used to study organizational aspects in their review of the effectiveness of judgmental adjustments of demand forecasts: "*This indicates the need for organization based studies that use interpretive research methods to establish, at a deep level, the beliefs and values of managers engaged in forecasting*" (p.18). More recently, Goodwin et al. (2018) have noted the lack of theoretical foundation on forecasting's organizational aspects in their forecast decisions review.

Notably, some researchers have made crucial attempts to respond to the call for the field's reviews. For instance, a seminal study in the field has identified communication gaps between forecast preparers and users as the main barrier to implementing forecasting techniques (Mahmoud et al., 1992). Similarly, Fildes & Hastings (1994) have identified a lack of awareness of statistical techniques and organizational design as impediments to forecasting activity. From a different angle, relevant studies have focused on the impact of the political dimension in the forecasting process, opening a stream of research that portrays more "naturalistic organizations" with power plays and attempts at influencing the

forecast to accommodate individual incentives (Bretschneider et al., 1989; Deschamps, 2004; Galbraith & Merrill, 1996; Lawrence et al., 2000). More recent contributions have revolved around appropriate organizational structure design, i.e., who owns the forecasting process (Protzner & Velde, 2015) and using incentives such as accuracy penalties to improve accuracy via behavioral alignment (Scheele et al., 2018). The few studies focused on processes in organizational forecasting literature deserve special attention. Davis & Mentzer (2007) have adopted a process approach conceptualizing forecasting as a dynamic capability embedded in a particular organizational climate and adjusted by feedback loops. Similarly, Oliva & Watson (2009) and Oliva & Watson (2011) have used case studies to demonstrate that process design can mitigate the effects of functional inclinations. The latter line of inquiry makes a more explicit reference to the interactions between team members and consensus reaching; however, the micro-processes that unveil the mechanisms at play are hinted at but not fully addressed.

Despite these efforts, the perception of a significant gap in understanding forecasting from an organizational perspective remains (Goodwin et al., 2018). This persistence might indicate an underlying—and largely inadvertent—paradigmatic entrenchment, partially engendered by the type of training and background of the researchers in the field, overweighted in the objectivistic-functional approach intrinsic to fields like statistics, operations, and econometrics (Fildes et al., 2003). The latter is relevant because the organizational aspects of forecasting cannot be studied exhaustively without relaxing the objectivistic presumption of the existence of a concrete reality "out there" that forecasters can accurately predict, neglecting the inherent subjectivity of some aspects of the forecasting process (Hogarth & Makridakis, 1981). Arguably, a relaxation of the concrete world assumption would increase the infusion of relevant organizational behavior and decision-making theory in the forecasting field, which could move the conversation into a "transition zone" between the dominant functionalist paradigm and interpretivist approaches (Gioia

& Pitre, 1990), where reality is observed through the eyes of the participants (Burrell & Morgan, 1979).

Dealing with subjectivity is not desirable but necessary when organizational dynamics—i.e., informational ambiguity, longitudinal contexts, behavior-bending incentives, the repetitive nature of some decisions, and hierarchical structures (Shapira, 2008)—are internalized as inherent characteristics of social interchange. Importantly, those organizational factors render accuracy a partial goal, especially when predictions anticipate a non-desirable outcome and require management intervention. In line with this reasoning, Einhorn & Hogarth (1981) have stressed that judgment accuracy (profit, errors, time) is conditional on context assumptions and a specific time frame, hence identifying "the conditional nature of optimality." When the context creates inevitable tension between goals—e.g., accuracy, motivation, and long-term viability—a decision-maker can only balance trade-offs based on subjective preferences (Hogarth & Makridakis, 1981).

Rosemberg (2023) found that these ideas have some empirical support; for instance, Sutcliffe & Weber (2003) have found a U-shaped curve relation between perceptual accuracy and firm performance. Unexpectedly, "humbly optimistic" managers—rather than accurate ones—improved performance using interpretation frameworks that mobilized action. In this case, inaccuracy propelled organizations to pursue goals that otherwise would be deemed unattainable:

Misperceptions may be beneficial if they enable managers to overcome inertial tendencies and propel them to pursue goals that might look unattainable in environments assessed in utter objectivity. Because environments aren't seen accurately, managers may undertake potentially difficult courses of action with the enthusiasm, effort, and self-confidence necessary to bring about success. Having an accurate environmental map maybe less important than having some map that brings order to the world and prompts action. (Sutcliffe, 1994; p.1374).

Occasionally, some forecasting researchers have recognized the multi-goal nature of the forecasting activity beyond accuracy. For instance, some researchers have suggested that organizations should judge the utility of sales forecasting on the extent to which it supports improved business performance metrics such as inventory levels, profitability, supply chain costs, and customer service (Mahmoud et al., 1992; Mentzer et al., 1999; Moon et al., 2003). Whereas McCarthy et al. (2006) reported that firms associated sales forecasting performance with business performance metrics like inventory levels (48%), customer service (30%), and supply chain costs (19%). However, these observations did not engender theoretical developments. In part because thinking about the forecasting process as an organizational instrument to influence outcomes rather than merely predict the future is not appropriate in a functionalist-objectivistic frame but in a world in which beliefs—at least partially—could shape reality; making forecasting, decision-making, and enactment blend under the intentionality of actors through self-fulfilling prophecies or self-altering predictions (Bolger & Wright, 1994; Einhorn & Hogarth, 1981; Henshel, 1993).

2. THE EFFECTS OF FUNCTIONALISM IN FORECASTING RESEARCH

"Your horse is a zebra, and zebras can't be tamed" (McGahan, 2007, p. 2)

Organizational theory in the forecasting context has been deemed "trendy, theoretical and complex" (Fildes et al., 2003). Notably, Rosemberg (2023) identified that these "paradigmatic blinkers"—imposing an almost exclusive focus on accuracy—might have attenuated the potential of two streams of research in the forecasting field: extracting forecasts from groups and learning from outcome feedback.

2.1 Extracting forecasts from groups

Research about extracting forecasts from groups has mainly focused on the mechanics—and risks—of forecast aggregation. Still, it has failed to describe the micro-processes that generate these outcomes, limiting its ability to make prescriptions. The inclusion of the biases and heuristics tradition in the

forecasting field—under the functionalist-objectivist paradigm—has almost exclusively highlighted individual judgment flaws that hamper accuracy (Eroglu & Croxton, 2010; Flyvbjerg, 2008; Kahneman & Lovallo, 1993; Lawrence & O’connor, 1995; Lee & Siemsen, 2016), where the collective aspect has emphasized the dangers of groupthink and bandwagon effects (Janis, 1982; Mellers et al., 2014; Sniezek, 1989; Surowiecki, 2005), and the potential bias accentuation effect of the group (Buehler et al., 2005). Consequently, the arsenal of the heuristics and biases tradition has cast a shadow of doubt on human judgment, especially when compared to algorithms (Fildes et al., 2009; Lawrence et al., 2006; Stewart, 2001).

However, this approach overlooks that the heuristics and biases program's objective was to generate a reliable descriptive model of human behavior so that researchers could design prescriptive models to improve judgment and *"not to come up with a list of biases."* (Shapira, 2008; p.4). Additionally, the literature often misses that the program's hypotheses have been tested mainly in laboratory studies focused on Individual's judgment against the outcome of normative models (Meehl, 1954; Tversky & Kahneman, 1974); hence, the validity of these findings in naturalistic settings is undergoing (Klein, 2008; Mosier et al., 2018). Furthermore, the forecasting literature has devoted insufficient attention to the observation that superficial information search and processing are the main drivers of error in human judgment (Slovic et al., 1977; Slovic & Lichtenstein, 1971) and that those inclinations could be mitigated and manipulated by specific group dynamics: sampling information from many sources, looking for disconfirming information, and using roles as "devil's advocates" (Hogarth & Makridakis, 1981; Mellers et al., 2014; Sunstein, 2015; Tetlock & Gardner, 2015), especially in teams that share mental models and are cohesive (Kerr & Tindale, 2004); hence, posing team interaction as the central dynamic in judgment within groups.

In one notable exception to the literature's neglect of team processes, Oliva & Watson (2009) and Oliva & Watson (2011) elevate the importance of teams' interactions within ritualized forums, e.g., "forecast consensus meetings." These

forums aim to forecast alignment via information sharing, anchoring discussions on an initial estimate, and open feedback, elements that could serve as group debiasing mechanisms and circumvent groupthink risks (Janis, 1982; Mellers et al., 2014; Surowiecki, 2005).

2.2 Learning from feedback

The second forecasting research stream restricted by functionalism has been learning from feedback. Arguably, at least part of the observed repetitive forecasting gaps could be justified by the intrinsic difficulty of predicting the future. For instance, researchers have asserted that uncertainty levels determine an event's predictability (Makridakis et al., 2010), while other views have suggested the existence of predictability limits related to the time horizon of the forecasts (Hogarth & Makridakis, 1981) or the intrinsic characteristics of the tasks, e.g., "task learnability" (Bolger & Wright, 1994). However, forecasting deviations' recurrence highlights organizations' difficulties in consistently learning from mistakes (Edmondson, 1996; Finkelstein, 2004; Madsen & Desai, 2010). Cannon & Edmondson (2005) have captured the latter concisely: *"The idea that people and the organizations in which they work should learn from failure has considerable popular support - and even seems obvious - yet organizations that systematically learn from failure are rare."* (p. 299)

Failing to learn from wrong forecasts has profound implications for the organizations' ability to plan and make decisions. Flyvbjerg (2014) refers to these implications in Mega Projects: *"With errors and biases of such magnitude in the forecasts that form the basis for business cases, cost-benefit analyses, and social and environmental impact assessments, such analyses will also, with a high degree of certainty, be strongly misleading."* (p. 9). Within this context, the scarce forecasting literature on feedback and learning has focused on the relative efficacy of different forms of information presentation: most recent error, feedback on biases, or forecast calibration (Bolger & Önköl-Atay, 2004; Goodwin et al., 2018; Lawrence et al., 2006). Additionally, some studies have described feedback limitations: point forecast feedback with no rationale or guidance does

not improve performance (Klayman, 1988), forecasters are not open to guidance as they overweight their rationale (Goodwin et al., 2011), and individuals do not accept suggestions but prefer to arrive at their conclusions independently (Parikh et al., 2001). In one exception, Legerstee & Franses (2014) presented a naturalistic study where forecasters received performance and cognitive process feedback, resulting in accuracy gains behind less frequent adjustments and making more downward interventions. However, the study was silent about the reasons for the improvement and studied forecasters individually with no interactions with their teams or contexts.

Notably, the organizational reasons that impede learning and adjustment from feedback remain unexplored in the forecasting literature; however, a deeper understanding of these causes can make interventions more apparent. For instance, understanding the unwillingness of individuals to discuss mistakes due to fear of penalties and individuals' instinctive tendency to deny or ignore errors to maintain self-esteem (Goleman, 1985; Sagan, 1995; Taylor, 1989) makes evident why singling out individuals diminishes the chances of reporting and learning from mistakes (Desai, 2014). These dynamics could also explain why sometimes the blame is placed on external factors, providing a plausible and non-confrontational cause that preserves harmony but impedes learning (Baumard & Starbuck, 2005). In addition to understanding the individuals' psychological reactions to mistakes, a focus on collective learning processes could help address agency issues—e.g., leaders maintaining the status quo—and forecasting game-playing (Mello, 2009; Meyer & Zucker, 1989). Research in this direction could also address collective attention issues (Hoffman & Ocasio, 2001; Ocasio, 1997), where shared mental models might filter out some events or make them go unnoticed, effectively becoming organizational blind spots.

Paradoxically, the quest for forecasting accuracy by refining techniques dilutes the potential of the study of organizational processes where accuracy becomes a byproduct of effective learning mechanisms from error and contexts that incentivize performance improvement over time (Mezias & Starbuck, 2003; Mezias & Starbuck, 2009). This new approach resembles the teachings in Zen in

the Art of Archery, where looking at the aim is replaced by a focus on the process of aiming, resulting in more chances to hit the mark eventually (Herrigel & Hull, 1953).

3. ORGANIZATIONAL NUANCES IN FORECASTING

"God gave all the easy problems to the physicists." – James March

Organizational aspects in a forecasting context might refer to different levels of analysis ranging from culture to structure design to processes. However, if interactions should be the primary consideration, focusing on small groups' dynamics might yield valuable insights since organizations have increasingly turned to work teams as a structural building block (Kozlowski & Ilgen, 2006). Hence, Rosemberg (2023) identified that prospective theoretical frameworks should acknowledge and integrate nuanced aspects of team interactions' content and form: learning from success versus failure, deliberation about performance, and the leader's role.

3.1 Learning from success versus failure

Even though failure is ubiquitous and inescapable in organizational life, the organizational learning literature, in general, and knowledge management, in particular, focus primarily on successful cases and best practices transfer. Researchers have focused on best practices transfer impediments (Alavi & Leidner, 1999; Szulanski, 1996), their codification and articulation (Nonaka, 1994; Zollo & Winter, 2002), and have elevated best practices transfer to the top priority of organizations' knowledge management efforts (O'Dell & Grayson, 1998; O'Dell, 2011). O'Dell & Grayson (1998) exemplify this view in a paper with the suggestive title of "If only we knew what we know: identification and transfer of internal best practices" where the fundamental underlying principle is that the knowledge needed for superior performance is embedded within the organization's success stories. The challenge is to uncover and actively share them. The origins of this approach can be traced back to Strategic Management views such as the firm's resource-based view (Barney, 1991; Penrose et al.,

2009), where knowledge is an asset that can generate competitive advantage. This view later evolved into the knowledge-based view of the firm, an outgrowth of the resource-based view that suggests that intellectual resources are the most strategically important of the firm's resources (Galbraith, 1974; Grant, 1996). Consequently, firms able to effectively manage and transfer these knowledge resources through best practices can expect to improve many aspects of firm performance, including customer service, production and structural costs, and innovation (Teece, 2004).

However, learning primarily from successes can be detrimental to long-term viability. (Miller, 1994) identifies risks of an excessive focus on success cases: a) inertial pressures in many aspects of the structure and decision-making process, b) extreme process orientation to replicate previous success, c) inattention translated into reduced intelligence gathering, and d) insularity translated in failure to recognize changes in the environment. Similarly, Baumard & Starbuck (2005) conclude that firms excessively focused on reproducing their successes create “behavioral programs” to make replication efficient. Once a particular success is being exploited, there are decreasing incentives to invest in information gathering, and firms might even try to block environmental changes or avoid analyzing errors. This strategy cannot be extended over time. They depict this behavior in a study of a European telecommunications company where they observed this firm learned little from the issues they investigated as their managers typically linked failures to exogenous trends or outsiders' involvement

Consequently, researchers from different domains have reached conclusions favoring learning from failure as the key to long-term viability. For example, knowledge from failure tends to depreciate less rapidly than from successful experiences (Madsen & Desai, 2010). Similarly, Kc et al. (2013) found that cardiac surgeons learn more from their successes than their failures but learn more effectively from others' failures than others' successes. These authors find that experiences with successes and failures tend to complement each other and motivate the most significant gains in future performance. In line with this, acknowledgment and processing of errors seem to be effective. Eggers & Song

(2014) report that when multiple entrepreneurs attribute their failures to outside factors and respond by simply changing industries for subsequent attempts, failure experience is harmful to the performance of future ventures. Conversely, when entrepreneurs attribute failures to factors under their control and modify these factors in subsequent ventures as a result of error analysis, failure experience enhances future venture performance.

If learning from failure benefits individuals and organizations, why do they fail to learn? The most apparent reason could result from the unwillingness of organizational members to discuss their involvement with failure due to fear of being penalized or used as a scapegoat (Sagan, 1995). Another plausible reason is psychological; individuals have an instinctive tendency to deny or ignore errors to maintain self-esteem and maintain an “illusion of control” over outcomes they deem essential (Goleman, 1985; Taylor, 1989). Related to both instances, the chances of reporting mistakes and learning from them diminish when the diagnostic method of failure is characterized by singling out individuals rather than focusing on issues related to processes or systems (Desai, 2014). Interestingly, sometimes the blame is attributed to external factors, significantly hampering the ability of the organization to learn from failure since a plausible and non-confrontational cause has been provided (Baumard & Starbuck, 2005). Another barrier could be agency issues; influential individuals might be willing to maintain the status quo even in the face of failure since introducing changes could affect their positions and result in a loss of influence (Meyer & Zucker, 1989). Other less apparent reasons could be linked to concepts such as threat-rigidity (Staw et al., 1981); where a significant failure produces a crisis that results in a stressful situation that reduces the search for new cues, eliciting habitual or default responses that perpetuate an unproductive cycle.

Finally, researchers have identified issues of collective attention as a cause for neglecting learning from failure (Hoffman & Ocasio, 2001; Ocasio, 1997). From this perspective, shared mental models might filter out some events or make them go unnoticed, effectively becoming organizational blind spots. Complementary to the abovementioned barriers, one additional element deserves attention as it is

relatively less covered in the literature: learning does not necessarily involve acting upon that knowledge. In their seminal paper, Einhorn & Hogarth (1981) have indirectly addressed that issue when differentiating judgment and actions. From this view, even if individuals within organizations arrive at the correct conclusions through learning, they might choose not to act if they perceive that acting upon that knowledge in a given context is risky.

The barriers to learning from failure are related to the characteristics of the organizational context and, hence, they could be addressed by understanding the impact on learning of constructs such as organizational culture (Pettigrew, 1979; Schein, 1990), organizational climate (Schneider et al., 2013; Schneider & Reichers, 1983) and organizational identity (Hatch & Schultz, 1997). Learning culture and learning climate are closely related and are usually assumed to significantly impact individual, team, and organizational learning (Alavi & Leidner, 2001; Gold et al., 2001). Linked to this, Janz & Prasarnphanich (2003) state that organizational culture is the most crucial input to organizational learning by encouraging or impeding knowledge creation and sharing. Therefore, an organization's culture should provide support and incentives and encourage knowledge-related activities by creating environments for knowledge exchange and accessibility (Davenport et al., 1998). Similarly, Edmondson (1996) concludes that learning cultures should generate positive "error climates."

Interestingly, her research on errors in a medical setting revealed that highly performing teams reported more errors, which allowed them to talk about errors, increasing error detection and correction. Even though these findings do not directly contradict the causal arrow of positive learning cultures allowing error disclosure and, ultimately, learning from error, they do open the question of whether the constant exposure to error processing generated positive aspects of learning cultures. One study proposed the need for the coexistence of a supporting learning environment, tangible learning processes, and leadership reinforcement (Garvin et al., 2008); even though the learning processes described are team debriefs, primarily targeted at discussing deviations, the authors do not develop further the role of error processing in generating positive

learning climates. This potential inverse relationship between error disclosure and discussion with positive learning climates implies a specific reduction of the harmful emotional impact errors (Keith & Frese, 2005). The latter is similar to the role of exposure therapy in inhibitory learning (Craske et al., 2014), where individuals reduce their fear of an object or an event when exposed to the same source of that fear, creating “fear habituation,” an effect that could reduce the anxiety of reporting and talking about mistakes, and “belief disconfirmation,” a process that could obsolete any preconceptions about the risk of making, discussing and reporting errors.

Intuitively, learning from best practices seems like a plausible strategy. However, in dynamic contexts, the tendency to focus on best practices could increase organizational inertia due to previous successful tasks' formalization (Baumard & Starbuck, 2005; Miller, 1994). More specifically, Miller (1994) associates an emphasis on learning from success with damaging outcomes: a) inertial pressures in decision-making, b) extreme procedural orientation to replicate previous successes, c) reduced intelligence gathering, and d) inadequate recognition of changes in the environment. Similarly, Baumard & Starbuck (2005) conclude that firms excessively focused on reproducing their successes create behavioral programs to make replication efficient, decreasing incentives to invest in information gathering and becoming *“less aware of events outside their immediate domains and less capable of diverse actions”* (p.283).

Contrary to the focus on success, the organizational learning literature provides empirical support favoring learning from failure as the key to long-term viability in different contexts such as the orbital vehicle launching industry (Madsen & Desai, 2010), cardiac surgery (Kc et al., 2013) and entrepreneurship (Eggers & Song, 2014). These authors find that experience with failures tends to yield the most significant gains in future performance, primarily when an error is acknowledged and processed.

In parallel, in the education field, comparative studies between Chinese and US teaching techniques have found that the former students systematically

outperform the latter in standardized math test scores. Stevenson & Stigler (1992) and Wang & Murphy (2004) explain this overperformance behind Chinese teachers' use of errors to prompt group discussion about mathematical concepts, promoting a classroom environment where students do not feel ashamed to make mistakes. Stevenson & Stigler (1992) linked these teaching practices to cultural beliefs: *"For Americans, errors tend to be interpreted as an indication of failure in learning the lesson. For Chinese and Japanese, they are an index of what still needs to be learned"* (p. 192). In this context, teachers and students *"dwell on errors,"* correcting them and then asking the students to explain their reasoning. Schleppenbach et al (2007) advance further in this direction in another comparative study and finds that the normalization of error reinforces the shared belief that it is expected to voice and discuss mistakes. They learned that discussing errors is so essential for Chinese teachers that they sometimes induce errors when starting a class conversation.

These findings in education have a certain parallelism with the conclusions of a study on nursing teams that revealed that the best-performing nursing teams reported more errors, were willing to discuss them, and received coaching from their superiors during these discussions (Edmondson, 1996). The study concluded that these teams had a better error climate, which increased error detection and correction.

3.2 Group deliberation about performance

Delphi is a collective forecasting method consisting of strictly separate forecasts from multiple individuals, revised using anonymous feedback from other Delphi panel members (Linstone & Turoff, 1975). In a review of Delphi's empirical studies, Rowe & Wright (2001) found that Delphi groups outperformed 'statistical' groups (aggregations of non-interacting individuals) in twelve of sixteen studies and outperformed standard interacting groups in five out of seven studies. Even though the causes behind this observed superior performance are not fully understood, one critical insight is that Delphi improves judgmental accuracy when outcome feedback includes "reasons" and not only statistics (Rowe & Wright,

1996). Paradoxically, even though Delphi is primarily described as a method of forecast aggregation, the interchange of reasons hints at the efficacy of group deliberation since this process could be described as deliberation "by turns" in which forecasters can alert each other about the use of unfitting framings, cognitive biases, and inappropriate cause-effect relationships (Goodwin & Wright, 2010).

Research on debriefs and performance appraisals provides additional support for the effectiveness of deliberation and performance feedback across different fields ranging from medicine (Gaba et al., 2001) to organizational training (Garvin et al., 2008). One meta-analysis, including 46 samples from 31 studies, concluded that debriefs improved effectiveness over a control group by approximately 25% on average (Tannenbaum & Cerasoli, 2013). Additionally, the authors defined specific debrief characteristics that seem to address the limitations observed in the forecasting literature. First, debriefs are active versus passive, translating into participation and self-discovery rather than being told how to improve; secondly, the primary intent should be developmental and non-punitive. Third, debriefs are centered on specific episodes or performance events, and finally, they include inputs from multiple team members and at least one external source.

One key observation is that team deliberation norms could have cultural implications on team interactions beyond effectiveness. For instance, the US Army After Action Reviews (AAR)—formalized retrospective analyses—have been linked to promoting team cohesiveness via group interaction, consensus building, intragroup competition reduction, and preservation of a positive atmosphere (Morrison & Meliza, 1999).

3.3 Team leaders as facilitators

One seminal study on best-performing nursing teams highlighted the impact of leadership on teams' ability to learn from mistakes (Edmondson, 1996). The study concluded that a leadership model that establishes a context for openly handling errors induces a shared perception of error normalization. This normalization triggers openness to report errors and generates discussions about

what went wrong, ultimately reducing errors in subsequent attempts. Conversely, the literature on group forecasting seldom integrates the fact that most forecasters operate in teams and small groups with a leader-staff configuration where voting systems are non-existent. However, the leader executes some form of compilation after receiving cues from the different members, a model of interaction termed "hierarchical decision-making teams with distributed expertise." (Hollenbeck et al., 1995).

Team leader-staff dynamics have been widely explored in the US Navy's Tactical Decision Making Under Stress program from 1990 to 1999 (TADMUS project). This project was one of the most extensive research efforts on team decision-making under stress, and it yielded essential findings regarding leadership's impact on team interactions and communication processes (Collyer & Malecki, 1998). TADMUS aimed at developing training plans to help leaders enhance team performance by setting team goals, managing conflict, ensuring participation, and observing group dynamics (Salas et al., 1998). Results showed that teams with leaders who had received formal leadership performed better in a series of tasks (Cannon-Bowers & Salas, 1998). Moreover, Tannenbaum et al. (1998) identified some effective leaders' behaviors and concluded that they acted as facilitators, contrary to the conception of leaders acting primarily as supervisors. These leaders engaged in self-critique to signal the team members it was "safe" to make and admit errors, focused on feedback on tasks instead of people to avoid affective conflict (but foster cognitive conflict), provided task-specific suggestions and ensured the discussion included teamwork feedback—team processes—instead of outcome feedback exclusively.

Kozlowski et al. (1996) attribute team effectiveness improvement behind facilitation to the triggering of collective metacognitive processing, especially when the "metacognitive musing" is verbalized as feedback about major task engagements and includes reflection upon performance and team processes. Since discussions alone do not facilitate shared cognition (Mathieu et al., 2000), these facilitation strategies could create shared mental models, i.e., team members agree at least on the group's objectives and the basic models used to

interpret reality. These mental models are one key element of deliberations' effectiveness since team members with dissimilar mental models hamper the team's performance (Mathieu et al., 2000; Smith-Jentsch et al., 2008).

4. MINDFUL ORGANIZING FRAMEWORK

"There is nothing more practical than a good theory" Kurt Lewin, 1952

Rosemberg (2023) analysis leads to a potential adaptive framework that includes some elements attenuated in the current forecasting literature: learning from error, group deliberation about performance, and the role of leaders as facilitators. For instance, he highlights that Gioia & Pitre (1990) have stressed that studying organizational dynamics such as power, conflict, and meaning negotiation cannot be analyzed thoroughly from a pure functionalist angle but needs a bridge to an interpretivist view. As an illustration of an interpretivist-functionalist bridge, they put forward the idea of structurationism, a circular process where agents' interactions enact structures that influence agents' behavior (Poole et al., 1985, 1996). The focus on actors' interactions highlights the importance of group communication that constitutes "micro-level decision-making interaction patterns" that form the building blocks of teams' processes (McGrath et al., 2000). Hence, contrary to the forecasting literature's approach, where interactions are neglected or seen as sources of biases (Fahimnia et al., 2019), relevant theoretical frameworks should incorporate social interaction dynamics, and their predictive power will depend on their grasp of these interactions' complexities (Poole et al., 1996).

Within this context, sensemaking could be classified as a structurationistic frame. The field's researchers have defined it as the perpetual process in which organizations try to interpret equivocal inputs and enact those interpretations back into the environment to make it more predictable (Starbuck, 2015; Weick, 1995a; Weick et al., 2005). Within that context, organizations engaging in sensemaking notice or perceive cues, create negotiated interpretations, and take deliberate action (Maitlis & Christianson, 2014). Crucially, these "sensemaking moves" are more nuanced than the mere exploitation of concrete cues in the

individual decision-making models, which collapses at the collective level when the politics of meaning are present (Patriotta, 2003).

One specific mode of implementing sensemaking is mindful organizing (Weick et al., 1999; Weick & Sutcliffe, 2007), a collective process consisting of learning mechanisms that focus attention, manage errors, and maintain alertness (Ocasio, 1997; van Dyck et al., 2005; Vogus & Sutcliffe, 2012); intrinsically ingrained in the interactions of groups (Vogus, 2011; Vogus & Sutcliffe, 2012).

The framework emerged from observations of teams in naturalistic, high-stakes contexts involving aircraft carriers, nuclear plants, and firefighting squads and described five effortful processes that create and sustain collective mindfulness, which improves decision-making (Weick et al., 1999; Weick & Sutcliffe, 2007). First, preoccupation with error describes the importance of actively reporting errors and treating all failures as opportunities to learn (Baumard & Starbuck, 2005; Edmondson, 1996, 1999; Kc et al., 2013; Madsen & Desai, 2010; Sagan, 1995), understanding the liabilities of success in the form of complacency (Sitkin, 1992), and inattention to weak cues (Miller, 1994). Second, reluctance to simplify interpretations manifests itself as a continuous search for different perspectives that helps individuals and groups counteract tendencies like confirmation biases and anchoring (Eroglu & Croxton, 2010; Kahneman & Lovallo, 1993; Lovallo et al., 2012; Lovallo & Kahneman, 2003), overreliance on experiential learning (Argote & Miron-Spektor, 2011), group cognitive centrality (Kameda et al., 1997), overreliance on old categories (Langer, 2000), and groupthink and bandwagon effects (Buehler et al., 2005; Janis, 1982). Third, sensitivity to operations entails high situational awareness, i.e., a "struggle for alertness" to catch errors in the moment, counteracting the challenges arising from overgeneralization of previous experiences (Cokely & Kelley, 2009; Lovallo & Kahneman, 2003) and superficial information search and processing (Fischhoff et al., 1977; Slovic & Lichtenstein, 1971). Fourth, commitment to resilience refers to believing in the imperfection of commonly accepted knowledge and the ability to bounce back from errors (Langer, 1998, 2000). Finally, under-specification of structures requires fluid decision-making in fast-paced environments, where team hierarchy

can be dissolved, so decision-makers change depending on the circumstances to overcome decision-making challenges like the "Deans decrease" effect (Bedeian, 2002) and hierarchical rigidity (Vaughan, 1996).

This framework's relevance for advancing forecasting from an organizational perspective lies in its counteraction of group judgment propensities reports in the literature; more specifically, three aspects are essential. Firstly, the five effortful processes of mindful organizing could illuminate the interactions that mitigate propensities towards biased and auto-pilot cognitive modes that hinder judgment and learning in forecasting contexts (Kahneman & Lovallo, 1993; Kahneman & Tversky, 1979; Lovallo & Kahneman, 2003; Tversky & Kahneman, 1974). Secondly, some mindful organizing processes, such as preoccupation with failure and reluctance to simplify interpretations, point to concrete social and cognitive mechanisms to process error as an essential antecedent to learning (Sitkin, 1992). Also, they facilitate the scrutiny of hypotheses to redefine previous categories (Langer, 2016). Finally, central to advancing forecasting from an organizational perspective, the thoroughly social aspect of mindful organizing leaves little room for heroic individuals, omnipresent leaders, or autonomous experts but requires "dense interrelations" (Weick & Roberts, 1993). These interrelations' content and form are essential mechanisms influencing outcomes such as reflective reframing, rigorous discussions of errors, or hypothesis questioning (Sutcliffe et al., 2016).

4.1 Debiasing group judgment with mindful organizing

"A tendency toward mindlessness is characterized by a style of mental functioning in which people follow recipes, impose old categories to classify what they see, act with some rigidity, operate on automatic pilot, and mislabel unfamiliar new contexts as familiar old ones." (Weick & Sutcliffe, 2007; p. 61)

The forecasting literature has emphasized the role of cognitive biases in diminishing the accuracy of judgment (Eroglu & Croxton, 2010; Flyvbjerg, 2008; Kahneman & Lovallo, 1993; Lawrence & O'connor, 1995; Lee & Siemsen, 2016),

where the collective aspect has emphasized the dangers of groupthink (Janis, 1982; Mellers et al., 2014; Snizek, 1989; Surowiecki, 2005) and the potential bias accentuation effect of the group (Buehler et al., 2005).

In parallel, the group decision-making literature has uncovered that judgment quality may suffer during group discussions. For instance, research on brainstorming has highlighted that group discussion tends to interfere with people's ability to immerse in a productive train of thought (Kerr & Tindale, 2004; Nijstad & Stroebe, 2006). This line of research has also highlighted other factors that yield "process losses" during group deliberation techniques such as brainstorming: an unwillingness to contribute ideas because of evaluation apprehension, and convergence via social comparison on a relatively low standard of performance (Larey & Paulus, 1999; Mullen et al., 1991). The "cognitive centrality" effect could explain this convergence, i.e., groups are more likely to choose ideas with the most significant overlap (Kameda et al., 1997), creating a convergence towards what is already known or believed, effectively working as an instance of collective confirmation bias (Bossche et al., 2011). For instance, In Tindale (1993) study, discussions involving conjunction problems were videotaped. The findings revealed that groups often exchanged individual judgments. Interestingly, over 60% of the time, they would choose an undefended judgment from a member without further debate—provided that the individual preference aligned with a shared representation.

A sensemaking lens would characterize these effects as positive asymmetry, a cognitive tendency—institutionalized through repetition and structure—that reduces the chances that decision-makers will consider cues that predict issues (Levinthal & Rerup, 2006; Weick et al., 2005). In connection, the high-reliability organizations (HRO) literature offers many examples of situations—often with disastrous consequences—in which people accommodate, explain away, or normalize discrepant cues, often because people operate within contexts that inadvertently reduce attention and mindfulness (Dunbar & Garud, 2009; Vaughan, 1996; Weick, 1988).

If mindlessness seems to be the default state of group interaction—leading to biased judgment—and mindful organizing seems to counteract those propensities; one central question would be how teams activate mindful organizing during team interactions. One central proposition of this analysis is that collective mindfulness states mediate the relationship between ritualized team engagements and collective judgment accuracy (Figure 1). These engagements can be conceptualized as a group-level treatment focused on debiasing. Additionally, the extent to which these engagements induce mindfulness can be moderated by using deep information processing artifacts such as analogical reasoning, characterized by its effortful attempts at understanding the underlying structure of events, problems, and decisions.

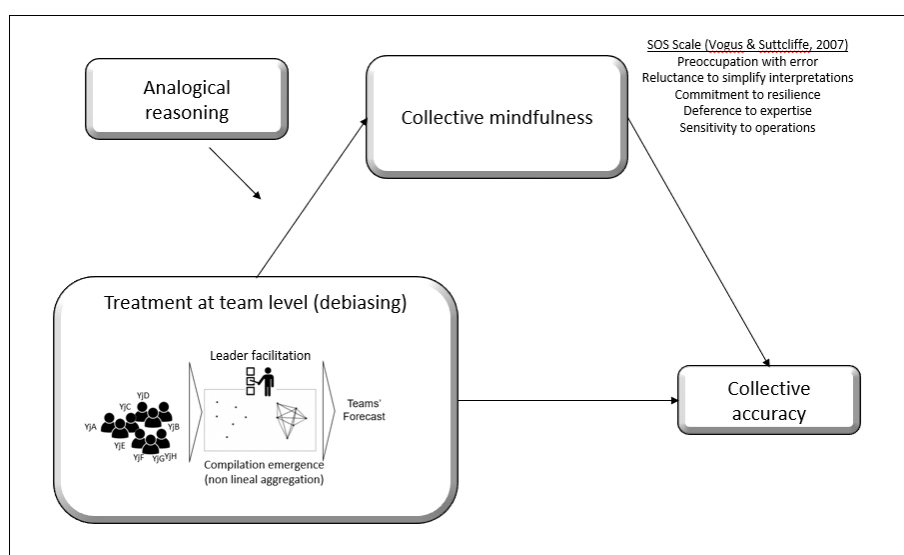


Figure 1. Theoretical model

4.2 Debiasing with a focus on episodic error

From a mindfulness organizing perspective, some elements might need to be present so that group interaction is adaptive in forecasting contexts: interactions need to be episodic, focus on dramatic events, and some degree of conflict—sustained for enough time—must be present.

Christianson et al. (2009) have conceived organizational learning as the revision of response repertoires triggered by errors that serve as audits—sometimes brutal—of existent repertoires. Importantly, these errors should generate

significant interruptions and exaggerated versions of the day-to-day stimulus to exert a change in organizing routines to justify the investment in effortful mindful engagements. For instance, Vogus & Colville (2017) have reported the beneficial effects of collective mindfulness only when a hospital nursing unit has a negative performance history of adverse events like harmful medication errors. Under those conditions, collective mindfulness acted as a problem-solving and emotion regulation resource. In contrast, collective mindfulness was associated with higher emotional exhaustion levels on units without a history of adverse events, depleting personal resources without tangible benefits. Similarly, Ray et al. (2011) have stated that team members will not question mindful organizing costs in high-reliability situations, such as nuclear power plants or aircraft carriers, because of the potentially fatal consequences. However, in other organizations, the constant pressures for efficiency could make it difficult to establish a relationship between organizational mindfulness and conventional financial performance measures unless managers recognize issues as critical, something more likely to happen if issues are dramatic or after major engagement incidents.

Conflict is another aspect of mindful encounters (Bossche et al., 2011). However, conflict only enhances team outcomes when it leads to deep-level processing, which allows a complete awareness of the complexity of the problems. Within this context, conflict—viewed as the adversarial review of information—should trigger "mindful scanning" (Vogus & Sutcliffe, 2012), which entails the exploration of the fringes of the current task (Fiol & O'Connor, 2003). Such practice could mitigate the temptation to fall into bandwagon effects and avoid the risks inherent to undisputed leaders' opinions, e.g., "deans decease" (Bedeian, 2002).

Finally, these team interactions must allow for sufficient time since new cues seem to impact decision-making when groups use enough time to reach consensus, avoiding premature closure risks. Dwelling on some aspects of the discussions is especially relevant when most members already share the same preference since people are perceived as more competent and credible when they share information that others already know (Kerr & Tindale, 2004). For example, in an experiment involving undergraduate students, they could only

overcome anchoring bias when researchers prompted them to think more carefully about their answers, so satisficing with the initial estimates was avoided (Epley & Gilovich, 2006).

4.3 Debiasing making use of analogical reasoning and reference classes

Researchers have stressed the fact that learning in organizations is impeded by the difficulty of creating, transferring, and absorbing knowledge (Argote, 1999; Nonaka, 1994; Szulanski, 1996), particularly when there is significant causal ambiguity between actions and performance in dynamic environments (Lippman & Rumelt, 1982). To reduce this ambiguity, Zollo & Winter (2002) have made a compelling case for departing from the dominant semiautomatic accumulation of experience towards more effortful and deliberate investments in articulation and codification of knowledge in the form of manuals, blueprints, and performance appraisals.

However, codification efforts have limitations and challenges. As the mindful organizing framework predicted, codification could increase organizational inertia due to previous successful tasks' formalization (Baumard & Starbuck, 2005; Miller, 1994). Moreover, one critical issue with the effectiveness of codified knowledge is the transfer of tacit knowledge, implicit know-how that is difficult to crystallize (Nonaka, 1994; Nonaka et al., 1994), given that people possess knowledge they cannot formally and explicitly explain (Polanyi, 2009). Furthermore, another limitation is linked to human cognitive processes since these manuals or tools could be manifestations of the mental models and preferences of individuals heavily influenced by biases such as confirmation, anchoring, planning fallacy, availability, over-optimism, recency, and underestimation of uncertainty (Kahneman & Lovallo, 1993; Lovallo & Kahneman, 2003; Tversky & Kahneman, 1974).

Knowledge codification seems like a plausible artifact to enhance collective learning and coincides with the mindful organizing framework on its requirements for sustained commitment to attentiveness (Levinthal & Rerup, 2006; Vogus, 2011); however, its limitations require consideration of how teams codify

information and use it in their interactions. Within these parameters, analogical reasoning could be an effective codification process.

Gick & Holyoak (1980) have demonstrated that analogies are potent devices to encapsulate and transfer knowledge from different domains. Their seminal study provided an experimental demonstration that a target problem can be solved using a similar situation from an unrelated domain when the relations between the analogy story and the target problem are mapped effectively. Moreover, Gick & Holyoak (1983) advanced this reasoning by empirically demonstrating that deriving a "problem schema"—a representation of the structural relationship from more than one analogy—enhanced knowledge transfer significantly since the use of several analogies can reduce the risk of mapping superficially the source and targets (Holyoak and Thagard 1995).

From a mindfulness perspective, Langer (1998) asserts that people can overcome biased judgment by the continuous creation and refinement of new categories and an implicit awareness of more than one perspective. Evidence exists that analogical thinking—via observation of analogous artifacts—can induce these mindfulness processes, allowing for deconstruction, symbolization, parody, and allegory (Barry & Meisiek, 2010; Davies, 2006).

The literature on analogies in management contexts is scarce but provides insights into how analogies can enhance judgment. Extracting knowledge from multiple analogies is embodied in the concept of "outside view," a deliberate effort to make use of analogical reasoning based on historical data to counteract the natural tendency of decision-makers to use an "inside view" (Kahneman & Lovallo, 1993): *"Decision makers have a strong tendency to consider problems as unique. They isolate the current choice from future opportunities and neglect the statistics of the past in evaluating current plans"* (p. 17). Similarly, Goodwin & Wright (2010) have reasoned that a critical aspect of taking an outside view—through a technique termed reference class—will be the forecasters' ability to use the correct analogies to filter out superficial similarities and retain a smaller number of cases with the most profound structural similarities. The effectiveness

of taking an "outside view" is supported empirically in different domains: effective competitive positions discovery (Gavetti et al., 2005), adjustments of stock valuation forecasts (Lovallo et al., 2012), geopolitical forecasts (Mellers et al., 2014), and making estimates about the costs of megaprojects (Flyvbjerg, 2008).

5. CONCLUSION

Mindful organizing can potentially redirect attention in the forecasting field towards more group-oriented research, given its intrinsic focus on group interactions about joint performance—e.g., preoccupation with error process—and the potentially adaptive dynamics it generates in the leader-member exchanges—e.g., reluctance to simplify specifications and underspecification of structures processes. In this sense, applying mindful organizing to forecasting teams could serve as a functionalism-interpretivism bridge, retaining the analytical power of the field directed towards accuracy while infusing the complexities of group interaction under a coherent framework.

Notably, recognizing that these interactions need to be episodic, structured around errors, possess an adversarial tone, and be facilitated by the leader provides an actionable template for creating and testing interventions that induce collective mindfulness, resulting in collective judgment debiasing and improving objective metrics like forecast accuracy. Thoroughly crafting interactions should shield these mindful encounters against adverse group decision-making effects described in the literature, such as bandwagon effects and the potential bias accentuation effect of the group.

Research in mindful organizing originated in the high reliability and safety literature and has spawned into other fields like medicine and education. However, to our knowledge, it does not exist in the forecasting field, partly because forecasting has not typically been approached as a task with high interdependence between different stakeholders, even though forecasting processes are vastly present in firms through structured multidisciplinary interactions such as Sales & Operations (S&OP) meetings and committees. Future studies complementing forecasting's accuracy pursuit (functionalism) with

interactionist approaches, such as mindful organizing, could shed light on the potential of forecasting as a bottom-up collective learning device, increasing the chances for the emergence of adaptive behavior at the level of the group that enhances performance.

III. EMPIRICAL CHAPTER

1. INTRODUCTION

One of the critical managerial challenges relates to forecasting future uncertainty. Forecasting effectively in organizational environments is intricate because it requires technical expertise to build predictive models and fit them to continuously changing task conditions while being cognizant of the *social context* (Fildes, 2006; Fildes & Hastings, 1994; Goodwin et al., 2018; Lawrence et al., 2006). Forecasts often represent several individuals' collective efforts, infusing social, cultural, and political differences that will likely affect the interpretation and communication of model outputs and, hence, the ability to produce action. Given the importance of the organizational environment embedding the forecasting process, it appears surprising that past research has primarily studied forecasting as *if* it represented an isolated, independent prediction task without implications for broader organizational decision-making (e.g., Lawrence et al. 2006). Consequently, building a comprehensive understanding of the drivers of judgmental forecasting performance remains elusive in many task domains. Beyond the frequently studied challenges associated with demand forecasting processes in organizations (Goodwin et al., 2018; Lawrence et al., 2006; Stewart, 2001), other forecasting domains—ranging from capital investment to I.T. project estimates and motion picture success—have also witnessed complexities that can result in poor predictive performance (Flyvbjerg, 2014; Kahneman & Lovallo, 1993; Lovallo et al., 2012). The persistence of low forecast accuracy in these domains has been a central concern in several past studies and industry reports. For instance, a large-scale survey of 5,400 IT projects with budgets above \$15M concluded that 45% of projects exceeded their budgets and achieved 56% less value than forecasted (Bloch et al., 2012).

Past research has highlighted the need for conducting further empirical studies on the role of organizational aspects in forecasting (Fildes, 2006; Fildes et al., 2009; Goodwin et al., 2018; Hogarth & Makridakis, 1981; Lawrence et al., 2006;

S. Makridakis & Wheelwright, 1977; Armstrong, 1988; Winklhofer et al., 1996). For instance, Hogarth & Makridakis (1981) highlighted the importance of evaluating forecasting processes through an organizational lens by considering their flexibility, underlying objectives, and level of control. Armstrong (1988) and Winklhofer et al. (1996) highlighted the need to improve our understanding of organizational challenges as a precondition to effective managerial forecasting. Years later, Lawrence et al. (2006) reviewed 25 years of judgmental forecasting literature and concluded that research in collective forecasting settings was largely neglected. Similarly, Fildes (2006) identified the lack of understanding of organizational issues as a critical gap in the literature, leading to a persistent discrepancy between forecasting theory and practice. From a methodological perspective, Fildes et al. (2009) highlighted *"the need for organization-based studies that use interpretive research methods to establish, at a deep level, the beliefs and values of managers engaged in forecasting"* (p.18). Finally, Goodwin et al. (2018) criticized the lack of a theoretical foundation to describe the relationship between organizational aspects and forecasting tasks.

Among the existing studies that contribute to a better understanding of the role of organizations in forecasting contexts, Fildes & Hastings (1994) recommended increasing managers' awareness of statistical forecasting techniques and improving communication between forecast preparers and users. Other studies focused on mitigating the effects of politics and power plays—which attempt to influence forecasts to accommodate individual incentives—by clarifying roles, increasing accountability, and debiasing forecasts (Bretschneider et al., 1989; Deschamps, 2004; Galbraith & Merrill, 1996; Lawrence et al., 2000). More recent contributions have examined the appropriateness of organizational structures and issues surrounding the ownership of the forecasting process (Protzner & Velde, 2015), as well as the use of incentives such as accuracy penalties to improve accuracy via behavioral alignment (Scheele et al., 2018). Finally, a few papers have employed organizational theory lenses to study forecasting performance. For instance, Davis & Mentzer (2007) have adopted a process approach conceptualizing forecasting as a dynamic capability embedded in a particular organizational climate and adjusted by feedback loops. Similarly, Oliva

& Watson (2009, 2011) have used a case study methodology to demonstrate that process design can mitigate functional inclination effects. The latter line of inquiry explicitly references interactions between team members and consensus reaching; however, the micro-processes that unveil the mechanisms at play are hinted at but not fully addressed.

Despite these efforts, the perception of a significant gap in understanding forecasting from an organizational perspective remains (Goodwin et al., 2018). This persistence might indicate an underlying—and largely inadvertent—paradigmatic entrenchment, partially engendered by the type of training and background of researchers in the field, overweighted in the objectivistic-functional approach intrinsic to fields like statistics, operations, and econometrics (Fildes et al., 2003). The latter is relevant because organizational aspects of forecasting cannot be studied exhaustively without considering the role of human interactions among the various stakeholders, which emerge from day-to-day communication and collaboration (Kozlowski & Chao, 2012; Kozlowski & Ilgen, 2016; Marks et al., 2000).

The primary objective of the present study is to empirically invigorate the theoretical link between forecasting and organizational decision-making by providing a framework to intervene in the communication micro-processes within forecasting teams. To do so, we adopt the perspective of mindful organizing to understand how interactions between forecasters influence their ability to discern emerging issues, manage errors, and maintain alertness (Ocasio, 1997; van Dyck et al., 2005; Vogus & Sutcliffe, 2012; Weick et al., 1999). Using an experimental design involving 40 hierarchical triads of forecasters—one leader and two team members—we systematically test the role of mindful organizing in judgmental forecasting in a controlled environment. Our key results reveal statistically significant improvements in forecast accuracy among groups that received interventions to foster interactions aligned with the mindful organizing framework. In particular, the regression analysis and robustness checks revealed that the experimental condition significantly improved accuracy by 24% vs. control groups measured by the Mean Absolute Percentage Error - MAPE (MAPE

leader with $p = 0.018$). This result suggests that the intervention led to more accurate decision-making, represented by the leaders' final forecasts. The study also found accuracy improvements in collective forecast aggregation metrics such as the group's MAPE and Mean Scaled Interval Score (MSIS)—without the final leader intervention. We found significant differences between the two groups regarding MAPE Avg Particip and MSIS4 (including four observations but excluding the time series peaks) but not MSIS6 (including all six observations). Finally, team communication analysis via chat functionality yields insights into more extended team communication (29% higher word count) and the leader's use of mindful organizing concepts and facilitating techniques in the intervention groups; both findings aligned with the expectations of the intervention.

The remaining sections of this chapter will unfold as follows. We provide an overview of extant research on organizational aspects of forecasting, such as obtaining forecasts from groups and using feedback while stressing potential oversights. We then present mindful organizing as an adaptive framework to address these challenges, specifically, its emphasis on collective attention to errors, systematic control of the quality of deliberations among team members, and reevaluating the leader's role within hierarchical forecasting teams from supervisor to facilitator of those team deliberations. Subsequently, we introduce our experimental design to test our central hypothesis and report the findings from an online experiment involving forecasting teams subjected to a mindful organizing intervention. In the final section, we discuss implications for the existing theory and practice of judgmental forecasting in organizational contexts.

2. THEORETICAL BACKGROUND

2.1. Group forecast extraction and task feedback

Eliciting group forecasts

Past research on organizational aspects of forecasting has often focused either on studying the performance of mechanical rules for aggregating individual judgments or on examining individual-level judgmental biases (Eroglu & Croxton, 2010; Flyvbjerg, 2008; Kahneman & Lovallo, 1993; Lawrence & O'Connor, 1995; Lee & Siemsen, 2016). At the same time, more general theories on group decision-making have highlighted the damaging effects of groupthink, herding, or bandwagon effects, which can accentuate existing collective biases (Janis, 1982; Mellers et al., 2014; Sniezek, 1989; Surowiecki, 2005). The implications of these effects for performance outcomes in the forecasting context remain to be explored.

Conversely, in the present study, we advocate for a return to the original view that errors in human judgment can often be attributed to superficial information search and processing (Slovic & Lichtenstein, 1971) and are not necessarily intrinsic, unmovable characteristics of human judgment as often presented in the heuristics and biases tradition (Shapira, 2008). We suggest that teams can—at least partially—compensate for those inclinations when using targeted group interventions that encourage the forecasting group to sample information from diverse sources, look for disconfirming information, and challenge their assumptions by using, for instance, group roles such as "devil's advocates" (Hogarth & Makridakis, 1981; Mellers et al., 2014). Among the few existing forecasting studies adopting this approach, Oliva & Watson (2009) examine the effectiveness of interactions between team members by focusing on the role of ritualized team forums, such as "forecast consensus meetings." Such meetings aim to align individual forecasts of team members through information sharing, discussions adjusting initial estimates, and providing open feedback, all of which

could debias collective judgment. However, previous research does not emphasize or describe group decision-making processes, such as the communication strategies used among team members (Kozlowski & Chao, 2012; Kozlowski & Ilgen, 2006).

The role of task feedback

Past studies have examined the efficacy of providing forecasters with different aspects of prediction performance, such as feedback on point forecast accuracy, the most recent errors, the presence of biases, or calibration issues, finding limited impact on accuracy performance (Bolger & Önköl-Atay, 2004; Goodwin et al., 2018; Klayman, 1988; Lawrence et al., 2006). Moreover, forecasters may be unwilling to accept external guidance and subsequently overweigh their rationale for making judgmental forecasts (Goodwin et al., 2011).

When studying forecasting in naturalistic settings, Legerstee & Franses (2014) focused on monthly SKU (Stock Keeping Unit)-level sales data in which individual forecasters received performance and cognitive process feedback. The study showed that forecast errors decreased when predictions were adjusted less frequently and downwards. Moreover, Athanasopoulos & Hyndman (2011) found that feedback improved predictive accuracy in the context of forecast competitions. However, both studies focused on feedback provided to individual forecasters and remained silent about the reasons that potentially underlie performance improvements.

Finally, research on the Delphi method—a method of forecast aggregation—has also addressed the role of feedback in forecasting (Linstone & Turoff, 1975). Interestingly, the Delphi method improves judgmental performance when outcome feedback contains qualitative information beyond purely statistical data (Rowe & Wright, 1996). Although Delphi represents a method of forecast aggregation, the collective exchange of qualitative reasons between panel members hints at the potential of group deliberation. Forecasters collaborating as teams can share relevant information, which increases information search and

processing, mitigating biases and the tendency to create stories behind weak cause-effect associations (Goodwin & Wright, 2010).

2.2. Team characteristics

Group deliberation about performance

Research outside the forecasting field on debriefs and appraisals suggests that such interventions can significantly improve group performance. Tannenbaum and Cerasoli (2013) reported that debriefs yielded a 25% improved effectiveness. The study identified some common characteristics of successful debriefs: members participated actively in the discussion, leading to self-discovery. Also, the tone was developmental versus punitive and centered on specific episodes or events. Evidence of the effectiveness of debriefing exists in different domains, ranging from health (Gaba et al., 2001) to organizational training (Garvin et al., 2008).

Team leaders as facilitators

Edmondson's (1996) seminal work on nursing teams emphasized the influence of leaders in the team's "error climate," a shared belief that errors are accepted as a regular occurrence in the learning process. This belief triggers openness to report errors and the discussion and exploration of possible root causes to reduce future errors. Conversely, in the forecasting literature, the team leader's existence and role are seldom discussed, and it is often assumed that the leader represents a rational agent effectively processing information cues from the diverse expertise of team members (Hollenbeck et al., 1995).

One crucial source of team leader-member interactions is the multiyear study commissioned by the U.S. Navy: Tactical Decision Making Under Stress program (TADMUS). TADMUS's goal was to enable leaders to improve team decision-making by following specific protocols for goal setting, conflict resolution, team engagement, and team participation by focusing on team interactions and, more specifically, on the team communication micro-processes (Collyer & Malecki,

1998; Salas et al., 1998). Results showed that teams with leaders who applied the adaptive protocols performed better across different domains (Cannon-Bowers & Salas, 1998). Moreover, Tannenbaum et al. (1998) identified that best-performing teams had leaders who intentionally facilitated team interactions instead of supervising them. For instance, these leaders were open to self-critique to instigate and normalize error discussion, ensured active team participation, avoided personalizing the responsibility of performance failures, and ensured feedback included team interactions and not only outcomes.

3. PROPOSED RESEARCH MODEL AND HYPOTHESES

This section will present a theoretical model of collective forecast accuracy improvement mediated by the mindful organizing construct. After that, an analysis of eliciting mindful organizing from the design and content of team engagement follows, and the study's central hypothesis is presented.

3.1. Debiasing group judgment with mindful organizing

"A tendency toward mindlessness is characterized by a style of mental functioning in which people follow recipes, impose old categories to classify what they see, act with some rigidity, operate on automatic pilot, and mislabel unfamiliar new contexts as familiar old ones." (Weick & Sutcliffe, 2007; p.61)

This study has asserted that the forecasting literature has emphasized the role of cognitive biases in diminishing the accuracy of judgment (Eroglu & Croxton, 2010; Flyvbjerg, 2008; Kahneman & Lovallo, 1993; Lawrence & O'connor, 1995; Lee & Siemsen, 2016). Moreover, studies of the collective aspect have warned about the dangers of groupthink (Janis, 1982; Mellers et al., 2014; Sniezek, 1989; Surowiecki, 2005) and the potential bias accentuation effect of the group (Buehler et al., 2005).

In parallel, the group decision-making literature has uncovered that judgment quality may suffer during group discussions. For instance, research on brainstorming has highlighted that group discussion interferes with people's ability to immerse in a productive train of thought (Kerr & Tindale, 2004; Nijstad & Stroebe, 2006). This line of research has also highlighted other factors that yield "process losses" during group deliberation techniques such as brainstorming, precisely, avoidance to contribute with new ideas, and a default convergence towards low-performance standards (Larey & Paulus, 1999; Mullen et al., 1991). The "cognitive centrality" effect could explain this convergence, i.e., groups are more likely to choose ideas with the most significant overlap (Kameda et al., 1997), creating a convergence towards what is already known or believed,

effectively working as an instance of collective confirmation bias (Bossche et al., 2011). For instance, Tindale (1993) analyzed videotaped discussions of conjunction problems and found groups typically exchange individual judgments and, more than 60% of the time, select an undefended judgment from one of the members without debate if the individual preference is plausible within a shared representation.

A sensemaking lens would characterize these effects as positive asymmetry, a cognitive default—established through repetition and structure—that decreases the odds that decision-makers will consider cues that predict issues (Levinthal & Rerup, 2006; Weick et al., 2005). In connection, the high-reliability organizations (HRO) literature offers many instances of circumstances—often with disastrous consequences—in which people disregard discrepant cues, often because people operate within contexts that inadvertently reduce attention and mindfulness (Dunbar & Garud, 2009; Vaughan, 1996; Weick, 1988). Therefore, mindful organizing is an effortful and rare collective state where people counteract their natural propensities, recognizing that default judgment might be unadaptive, but adaptive judgment is possible. Table 1 maps mindful organizing processes to group judgment propensities in the decision-making and forecasting literature; for instance, issues like complacency and inattention could be counteracted with a focus on errors, and cognitive centrality could be counteracted with a group reluctance to simplify interpretations.

One central proposition of this analysis is that collective mindfulness states affect ritualized team engagements, influencing collective judgment accuracy. Correspondingly, if mindlessness seems to be the default state of group interaction—leading to biased judgment—and mindful organizing seems to counteract those propensities, one central concern should involve the design of interventions that instill the emergence of mindful organizing during team interactions.

Table 1: Mindful organizing counteracts group judgment propensities

Mindful organizing processes	Group judgment propensities in forecasting and decision-making literature
Preoccupation with error	Complacency (Sitkin, 1992), extrapolation of past success (Baumard & Starbuck, 2005; Kc et al., 2013; Madsen & Desai, 2010), explaining away mistakes (Sagan, 1995), stigmatization of error (Edmondson, 1996, 1999), inattention to weak cues (Miller, 1994)
Reluctance to simplify interpretations	Confirmation biases and anchoring (Eroglu & Croxton, 2010; Kahneman & Lovallo, 1993; Lovallo et al., 2012), overreliance on experiential learning (Argote & Miron-Spektor, 2011), group cognitive centrality (Kameda et al., 1997), overreliance on old categories (Langer, 1998, 2000), groupthink and bandwagon effects (Buehler et al., 2005; Janis, 1982)
Sensitivity to operations	Overgeneralization of previous experiences (Cokely et al., 2018; Lovallo & Kahneman, 2003), superficial information search & processing (Slovic et al., 1977; Slovic & Lichtenstein, 1971)
Commitment to resilience	Failure as a definitive state, overreliance on old categories (Langer, 1998, 2000)
Underspecification of structures	"Deans decrease" effect (Bedeian, 2002), hierarchical rigidity (Vaughan, 1996)

3.2. Elicitation of mindful organizing: team engagement design

Team deliberation is essential to mindful organizing since language is the critical sensemaking artifact (Browning & Boudes, 2005; Klein et al., 2010; Weick et al., 1999). From a mindful organizing perspective, language is not a representation of cognition but the conceptual system itself (Lee & Schwarz, 2014; Patriotta, 2003); hence, debate and deliberation are not preferable but necessary to elicit mindfulness (Klein et al., 2010). Additionally, some elements seem to be required in these interactions to elicit a mindful response: episodic engagements about error, some degree of conflict—sustained for enough time—and a leader who acts as a facilitator.

Christianson et al. (2009) have conceived organizational learning as the review of response repertoires triggered by errors that serve as audits—sometimes brutal—of existent repertoires. Importantly, these errors should generate significant interruptions and exaggerated versions of the day-to-day stimulus to

exert a change in organizing routines to justify the investment in effortful, mindful engagements. For instance, Vogus & Colville (2017) reported the beneficial effects of collective mindfulness only when a nursing unit had a history of adverse experiences such as medication errors. In those circumstances, collective mindfulness was a resource for solving problems and regulating emotions. In contrast, collective mindfulness was related to higher emotional exhaustion levels in units without adverse track records, depleting personal resources without tangible benefits. Significantly, debating about errors instead of individual involvement (e.g., accidents, interruptions, and surprises) can trigger reflection and deep-level processing (Bossche et al., 2011; Ron et al., 2006).

The attention required by mindful organizing is costly and cannot be sustained continuously outside risky environments. Ray et al. (2011) have stated that teams will not question mindful organizing costs in high-reliability settings, such as nuclear power plants or aircraft carriers, because of the potentially fatal consequences. However, in other organizations, the pressures for efficiency could make mindfulness seem financially unviable unless managers recognize issues as critical, something more likely to happen if dramatic issues are linked to major engagements or novel situations. Hence, these engagements need to be episodic and perceived as necessary.

Another aspect of mindful encounters is the existence of debate. Some degree of conflict enhances team outcomes when it leads to deep-level processing, which allows a complete awareness of the complexity of the problems (Bossche et al., 2011). Within this context, conflict—viewed as the adversarial review of information—should trigger "mindful scanning" (Vogus & Sutcliffe, 2012), which entails the exploration of the fringes of the current task (Fiol & O'Connor, 2003). Such practice could mitigate the temptation to fall into bandwagon effects and avoid the risks inherent to undisputed leaders' opinions, e.g., "deans decease" (Bedeian, 2002). Dwelling on some aspects of the discussions is especially relevant when members share a preconceived preference since people are perceived as more competent when they repeat information known to the group (Kerr & Tindale, 2004). Also, conflict avoids premature closure risks because new

cues require time to impact the decision-making process. For example, in an experiment involving undergraduate students, anchoring bias was only overcome when individuals were instigated to reflect cautiously on their answers, so satisficing with the initial estimates was avoided (Epley & Gilovich, 2006).

The final element is the introduction of the team leader as the vehicle of mindful team interactions—and the focus of the intervention—since the most common configuration in teams is leader-staff, where there are no voting systems, but the leader has the final call (Hollenbeck et al., 1995). Research shows that leader-member exchange (LMX) can create contexts where mindful organizing can result (Morgeson & Hofmann, 1999), that senior managers can elicit broader scanning and questioning of interpretations (Fiol & O'Connor, 2003), and that best-performing teams have leaders who acted as facilitators in team engagements (Tannenbaum et al., 1998).

In this dissertation, we explore the impact of mindful organizing interventions on group forecast accuracy with two hypotheses. First, we hypothesize that implementing mindful organizing intervention—including elements such as leader discussion facilitation, team debate, and focus on error—will enhance the team's forecast accuracy compared to the control groups, represented by the accuracy of team leaders' final decision after facilitating the team discussion and accessing individual forecasts from the team. This first hypothesis emphasizes the observation that leaders are entitled to make final judgments in naturalistic hierarchical teams.

H1: Mindful organizing interventions will improve the team leader's forecast accuracy relative to the control group.

Our second hypothesis posits that mindful organizing interventions will influence team forecast aggregation accuracy. This premise complements the first hypothesis; if the leader has access to more accurate inputs from team members, the aggregation of the team members' judgments should also be more accurate. As team members engage in mindful processes, their ability to assess and

contribute accurate judgments will improve, ultimately enhancing the quality of judgmental aggregations.

H2: Mindful organizing interventions will improve the mean accuracy of group members' forecasts, resulting in more accurate judgmental aggregations relative to the control group.

4. METHODS

4.1. Task Design & Data Sample

We used an experimental study to investigate the impact of a group-level intervention—based on mindful organizing principles—on the forecasting performance of hierarchical triads. Importantly, this study uses newly formed teams following (Kozlowski & Chao, 2012) observation that the most critical phenomena have already emerged in previously formed teams. The experiment has two stages: a team deliberation phase, where members discuss and submit an individual forecast, and a leader forecast phase, where the leader submits a final judgment based on the previous phase's input. We compare the accuracy of intervention and control groups in both stages. Specifically for the deliberation phase—containing three individual forecasts—we create two aggregated measures using the group's single-point estimations: an arithmetic average and a group interval (Gneiting & Raftery, 2007).

Participants' recruitment was performed online via the Prolific platform. Prolific provides researchers with tools for ensuring the quality of participant data, such as by verifying participants' identities and checking for duplicates or suspicious behavior. It also ensures that the participants are fluent in English and have no history of cognitive or neurological disorders that could affect their ability to complete the forecasting task. Another Prolific advantage is its diverse and large participant pool, which can be accessed swiftly, allowing for easier recruitment and faster data collection.

The validity of using online subjects for experiments is a matter of debate in psychology and social sciences. While some researchers have expressed concerns about the quality of data collected from online participants, several studies have shown that online data can be as reliable and valid as data collected from traditional lab-based studies (Buhrmester et al., 2016; Crump et al., 2013; Gosling et al., 2004; Peer et al., 2017). Even though these studies concluded that

online participants could provide reliable and valid data, they caution researchers about specific steps—addressed in this study through pilots and tests—to ensure data quality:

- **Attention and engagement:** Online participants can easily distract or disengage during a study. Researchers must design their experiments to maintain participants' attention and engagement. In this study, each experiment phase, including intervention, control, and instructions, had comprehension and attention checks to avoid this issue. Additionally, the payout was dependent on successful completion.
- **Technical issues:** Online experiments are vulnerable to technical issues such as slow internet connections, browser compatibility issues, and software bugs. Researchers need to thoroughly test their experiments before launching them and be prepared to address technical issues as they arise. For example, initial pilots on this experiment identified that the interface was not mobile-friendly, limiting participants to use only personal computer devices.
- **Generalizability and sample bias:** Sears (1986) highlighted lab research's narrow focus on participants from undergraduate programs ranging from 18 to 23 years old and highly educated. Prolific allows for a wider age range, as shown in Table 2 (mean age of 39.2 years and std dev of 14.9 years), and more diverse educational background recruiting (17% of the sample were students and 83% were non-students).

Table 2: Sample age distribution

Sample age range	%
18-23	7%
24-40	56%
41-50	20%
50 and above	18%
Total	100%

- **Ethics and data privacy:** Prolific provides privacy notices meeting the standards of the European (General Data Protection Regulation) and U.K.

data protection laws. Additionally, Prolific uses encrypted HTTPS connections, secured by Transport Layer Security (TLS), and participants are anonymized using a unique participant I.D. (24-character alphanumeric).

4.2. Experimental Protocol

The experiment divided participants into 40 triads, each comprising two team members and one team leader. We recruited participants in real-time; once a triad was formed, the platform randomly allocated it to either the treatment or control conditions. Furthermore, participants were randomly assigned either the role of a "team leader" or a "team member."

In the treatment condition, groups received training in mindful organizing, emphasizing various techniques for facilitating group discussions (Morrison & Meliza, 1999; Tannenbaum et al., 1998; Tannenbaum & Cerasoli, 2013; Weick et al., 1999). This training phase comprised four concise videos (between 90 and 120 seconds each) and several comprehension check questions that ensured the principles of mindful organizing were understood. The author recorded these videos by creating and merging Microsoft PowerPoint slides with an audio track containing a narration with the author's voice.

Regarding the content of the treatment condition videos, the first highlighted the advantages of leaders as facilitators in team engagements, while the second video described facilitation techniques such as encouraging participation, discussing team dynamics, and connecting the current task to prior performance (Kozlowski & Ilgen, 2006; Tannenbaum et al., 1998). The third video provided questions directly linked to the five processes of mindful organizing (Weick et al., 1999): "What did we assume that did not happen?" "What are we missing?" "Are we fully immersed in the task?" "What actions can be adjusted and tested in the next round?" and "Is someone in the team better at this task, and should we let this person decide?" The comprehension check questions following each of the three videos provided a list of four options, whereby two of the answers were

correct. A fourth and final summary video summarized the content of the previous videos.

Separately, participants in the control group received decoy training, which consisted of a video about transformational leadership from YouTube.com (SkillsPacks, 2015). The key objective of the decoy video was to ensure that all participants in the experiment experienced a similar setup in terms of structure and timing and to avoid participants assuming their group assignment, i.e., blinding (Juul et al., 2021). Transformational leadership enhances organizational performance through leader behaviors such as motivation, inspiration, individualized consideration, and intellectual stimulation (Burns, 1978). This concept was selected as a decoy because it influences organizational performance via organizational culture changes, which requires lengthy behavioral shifts and trust-building processes (Bass & Avolio, 1993; Bass & Riggio, 2005; Wang et al., 2011). These effects are unlikely to impact the performance of this short-term experiment task. Similarly to the intervention groups, participants could only progress to the actual forecasting task after answering the comprehension checks correctly.

After the training phase, the actual task involved a multi-round team forecasting exercise that lasted, on average, 27 minutes across six rounds during which forecasts were elicited. All teams observed the same time series of monthly sales data with 48 points as input for the forecasting task (Makridakis & Hibon, 2000). In each round, their task was to forecast the next value of the time series (i.e., point 49 in round 1, point 50 in round 2, point 51 in round 3, and so forth). Six rounds were selected for two reasons: a) keeping the exercise close to 30 minutes completion as drop rates increased after that in the pilot tests, and b) provide enough repetition for the intervention to emerge; for instance, Hesser (2015) provides a general recommendation of a minimum of four task attempts when investigating internet interventions.

Each round consisted of two steps (Figure 2); in the first step, participants had access to a "deliberation screen" where they could see the time series and,

crucially, use a chat function to discuss and exchange points of view regarding the forecasting task. Moreover, regardless of whether the assigned role was the team leader or team member, participants had access to a submission area where they were required to submit an individual forecast judgment. In step two, the team leader was provided exclusive access to a "leader screen," where she could see a summary of the team members' submissions and make a final judgment on behalf of the group. This process went on for six rounds. Importantly, participants could see the team leader's final forecast submitted in the previous round and the actual demand values on the discussion screen to create a feedback loop and elicit discussion. After completing all six rounds, participants completed two surveys, one Big Five personality test (Rammstedt & John, 2007), and the SOS mindful organizing inventory (Vogus & Sutcliffe, 2007).

In our analysis of the forecasting data, we use the data in two ways. First, we use the team leader's forecasts—submitted on behalf of the group from the "leader screen"—to compute team accuracy. Secondly, we use the forecasts submitted in the "deliberation screen" individually to compute forecast aggregation metrics. To incentivize participants, we awarded them a flat payment of £5 in return for completing the task. In addition, participants could earn an additional, accuracy-based bonus of up to £5, calculated as $£5 \times (1 - \text{Mean Absolute Percentage Error [MAPE]})$. This MAPE calculation used the leader's six forecasts (from the "leader screen"); hence, the same group members received the same bonus.

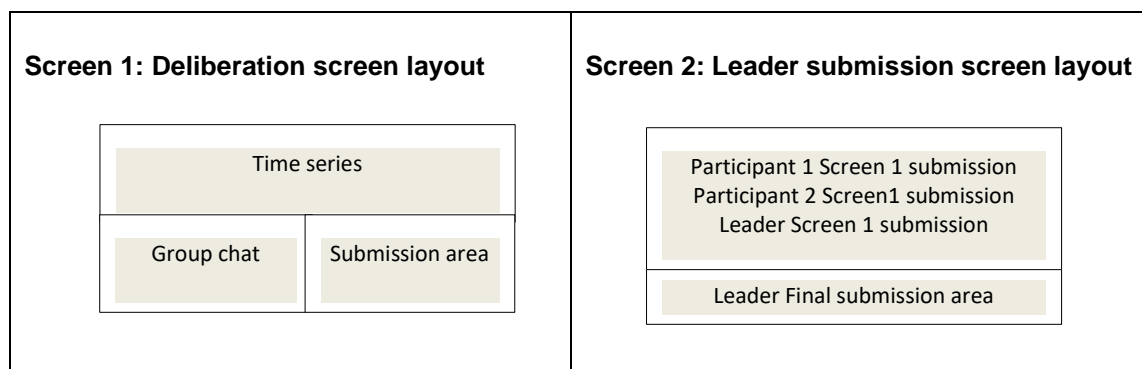


Figure 2: Forecasting exercise screen layouts

4.2.1. Time series data generation

Our experiment relies on the data generation procedure outlined in Kremer et al. (2011). In particular, a non-seasonal artificial times series for demand (D_t) was created to control the noisiness of the forecasting task employed in the experiment. The model manipulates two key characteristics of the time series data with two independent random variables: ε_t with its parameter n creates temporary shocks and v_t with its parameter c creates permanent shocks. By adjusting the values of c and n , different types of demand environments can be created, such that when n is set to zero, the model produces purely random walks, whereas setting c to 0 results in stationary "white noise." Between these extremes, increasing the value of c (standard deviation of permanent shocks) and decreasing the value of n (standard deviation of temporary shocks) generally lead to higher overall variability in the demand process. This combination introduces more persistent shocks that impact subsequent periods, resulting in more pronounced fluctuations. In the present study, we used n and c values equal to 40 to balance noise and constant level changes, hence the exercise difficulty. Figure 2 shows the time series generated by the model.

$$D_t = \mu_t + \varepsilon_t', \text{ where } \varepsilon_t \sim N(0, n^2)$$

$$\mu_t = \mu_{t-1} + v_t', \text{ where } v_t \sim N(0, c^2)$$

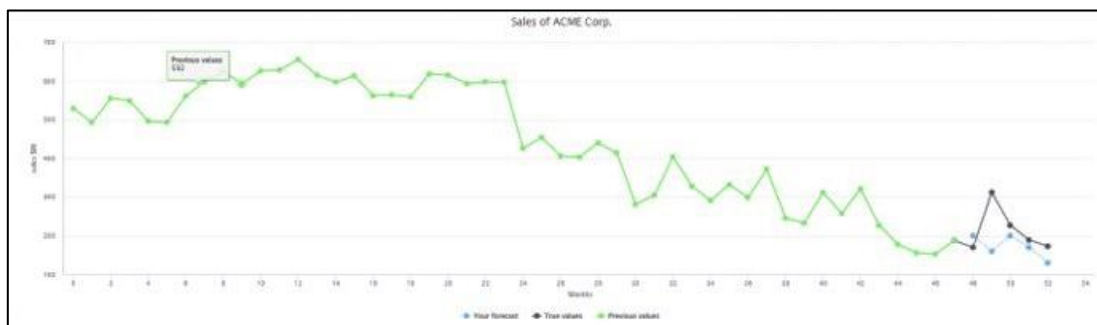


Figure 3: Time series including team forecasts and real values (last five values)

4.3. Operationalization of Key Constructs

Dependent variables: measures of group accuracy

The premise of this study is to determine whether a group intervention focused on the team leader's behavior will impact overall team forecasting performance using a generated time series (Figure 3). The study collects forecasting judgments from individual team members during the group deliberation phase and a final team forecast submitted by the assigned leader (Table 3). Therefore, the primary unit of analysis focuses on team leaders' final forecasting judgment performance in each of the six rounds using Mean Absolute Percentage Error (MAPE). In addition, we measure the team forecasting performance using aggregations: MAPE of the individual team members—as a simple arithmetic average for each round—and their interval accuracy using the Mean Scaled Interval Score (MSIS) (Gneiting & Raftery, 2007).

In particular, each team's forecast interval considers the highest and lowest forecasting judgments provided by the team members. An interval score is calculated using the team's upper and lower bounds (p_u and p_l) compared with the actual values (y) for a given significance level (α), the indicator function 1 is 1 when the condition holds; otherwise, it is 0. Each of the six rounds yields an average score, which is then scaled by dividing by the Mean Absolute Error (MAE) to yield the total task score.

$$\text{Interval score each round} = (p_u - p_l) + \frac{2}{\alpha} (p_l - y) 1(y < p_l) + \frac{2}{\alpha} (y - p_u) 1(y > p_u)$$

Table 3: Measures of accuracy

Step	Collects forecast from	Used as input for
Step 1: Deliberation Screen	Team Member 1	<i>Hypothesis 2 testing:</i> <i>MAPE group average,</i> <i>Interval accuracy (MSIS)</i>
	Team Member 2	
	Team Leader	
Step 2: Leader screen	Team Leader	<i>Hypothesis 1 testing:</i> <i>MAPE leader</i>

Independent variable: experimental treatment versus control group

Given the between-subject design of the experiment, our key independent variable will be a binary indicator, which takes a value of 1 if the participant has received the mindful organizing intervention prior to the task and a value of zero if she completes the forecasting task without prior training.

Manipulation check: mindful organizing scale

Several validated scales of mindful organizing are available in survey measures (Sutcliffe et al., 2016). This study will use a one-dimensional instrument called the Safety Organizing Scale (SOS, Figure 4), later called the Mindful Organizing Scale, which consists of a nine-item survey (Vogus & Sutcliffe, 2007). The scale shows reliability and several validity measures, such as convergent, criterion, and discriminant validity. Also, additional confirmatory factor analyses and several validity tests—e.g., internal consistency, criterion, and nomological networks—have been performed when translating the instrument to German, French, Italian, and Spanish (Ausserhofer et al., 2013; Renecke et al., 2020). The scale has been validated across many domains, including nuclear power plants and nursing environments (Renecke et al., 2020; Sutcliffe et al., 2016). In addition to the validity considerations, the SOS instrument allows for group-level aggregation, a critical design feature for this experiment (Vogus et al., 2014; Vogus & Sutcliffe, 2007). SOS collects individual responses on its nine items but aggregates the values across all respondents (each item refers to “we” to emphasize its collective capability nature).

Finally, the SOS scale has nine items versus more than forty in other cases. One of them is the instrument presented by Ray et al. (2011), a multifactor measure that validated the five mindful organizing processes using a 42-item survey. For this study, parsimony is critical to avoid overwhelming the participants and increasing drop rates since online participants are more prone to quitting due to lower sunk costs of participation, e.g., no need for commuting to the experiment location, and lower social cost, e.g., online anonymity (Zhou & Fishbach, 2016).

Concept ²⁴⁻²⁶	Definition ²⁴⁻²⁶	SOS Survey Item(s)
Preoccupation with failure	Operating with a chronic wariness of the possibility of unexpected events that may jeopardize safety by engaging in proactive and preemptive analysis and discussion	When giving report to an oncoming nurse, we usually discuss what to look out for We spend time identifying activities we do not want to go wrong
Reluctance to simplify interpretations	Taking deliberate steps to question assumptions and received wisdom to create a more complete and nuanced picture of ongoing operations	We discuss alternatives as to how to go about our normal work activities
Sensitivity to operations	Ongoing interaction and information-sharing about the human and organizational factors that determine the safety of a system as a whole	We have a good “map” of each other’s talents and skills We discuss our unique skills with each other so we know who on the unit has relevant specialized skills and knowledge
Commitment to resilience	Developing capabilities to detect, contain, and bounce back from errors that have already occurred, but before they worsen and cause more serious harm	We talk about mistakes and ways to learn from them When errors happen, we discuss how we could have prevented them
Deference to expertise	During high-tempo times (ie, when attempting to resolve a problem or crisis), decision-making authority migrates to the person or people with the most expertise with the problem at hand, regardless of their rank	When attempting to resolve a problem, we take advantage of the unique skills of our colleagues When a patient crisis occurs, we rapidly pool our collective expertise to attempt to resolve it

SOS indicates Safety Organizing Scale.

Figure 4: SOS items and correspondence with Mindful Organizing from Vogus & Sutcliffe (2007)

Control variables: Big Five personality test

Some studies have explored the influence of individual differences in personality traits on judgmental accuracy, albeit with mixed results. For instance, Poore et al. (2004) found aptitude to be a stronger predictor of predictive performance than Big Five personality traits, which were only weakly correlated— except for conscientiousness— with Brier accuracy scores. In contrast, Eroglu and Sanders (2021) found that forecast adjustments positively affected average accuracy, and personality traits such as openness to experience correlated with better individual accuracy. In more general cognitive domains, a meta-analysis of the Big Five traits found conscientiousness as the more robust predictor of general cognitive performance (Trapmann et al., 2007). Hence, considering the five personality traits, the present experiment focused on controlling for individual differences in forecasters' conscientiousness and openness, which we operationalized using Rammstedt & John (2007) 10-item version of the Big Five instrument, which simplifies the original 44-item instrument (John et al., 1991). The same parsimony considerations apply in this case as in the SOS Scale.

5. DATA ANALYSIS AND RESULTS

Our sample included 120 participants grouped into 40 triads of forecasters. Regarding demographics, 83% of participants were U.K. nationals, and 8% were from the U.S. Additionally, 93% were older than 23 (38% older than 40), and 17% were current students (student or non-student were the only occupation options). Regarding gender, 65% of the participants were male and 35% female. Finally, 83% were of Caucasian origin. The average performance-based pay was £3.5 (on top of the base pay of £5), and a triad's average task completion time was 27 minutes. Analyses were performed using the IBM SPSS software version 28.0.1.1, and graphs were generated using Datatab.

5.1. Manipulation check

To test whether the mindful organizing intervention was successful, we investigated whether forecasters in the treatment group exhibited significant differences regarding their responses to the SOS Mindful Organizing instrument relative to participants in the control group. The SOS scale measures how much a triad has internalized the mindfulness organizing principles. Figure 5 shows the distribution of the SOS Mindful Organizing instrument responses for both control and intervention groups, with the manipulation condition showing higher mean SOS mindful scores than the treatment group (control mean: 2.33, treatment mean: 2.98).

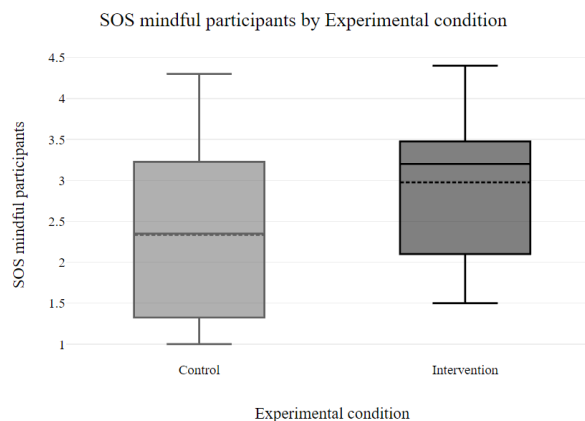


Figure 5: SOS in control and intervention groups

Two outliers were eliminated, corresponding to one triad with a participant who did not respond and one with a participant who rated the max scale value in all answers (implausible). Normality tests suggest that the data does not significantly deviate from a normal distribution with p-values higher than 0.05 (Kolmogorov-Smirnov=0.19, p-value= 0.466 and Shapiro-Wilk=0.91, p-value=0.094). Two tests were conducted. A one-tailed t-test was performed since the intervention was expected to improve the instrument outcomes in one direction, resulting in a significant difference in the mean values between control and treatment groups ($t=-1.98$, p-value = 0.028). Additionally, a 2-tailed test found a p-value of 0.055. These results suggest that our mindful organizing intervention was likely to be successful.

5.2. Measuring team accuracy using leaders' judgment

We start the main analysis by comparing the accuracy of the forecasts between control and intervention groups by measuring the leader's final forecast Mean Average Percentage Error (MAPE leader, step 2-leader screen). We eliminated three observations at 99% of the distribution (two in the control group and one in the intervention group), all with MAPEs above 54%, which is considered very weak (Lewis, 1982). In these cases, the leader input implausible values like 0 or 1, which inflated the MAPE calculations. MAPE leader showed a mean of 25.1% in the intervention condition vs. 31.0% in the control groups (24% more accurate).

Before performing a t-test, normality assumptions were checked, suggesting the data does not significantly deviate from normality (Kolmogorov-Smirnov=0.16, p-value=0.67 and Shapiro-Wilk=0.91, p-value=0.096). The result of a two-tailed t-test indicates that this difference is statistically significant, with a p-value of 0.016 (table 4).

Table 4: Experimental conditions accuracy

		N	Mean	Std. Deviation	Std. Error Mean
MAPE leader	Control	18	31	7.74	1.82
	Intervention	19	25.11	6.37	1.46
t-test			t	df	p (2-tailed)
MAPE leader			2.54	35	.016

Additionally, we used a regression analysis to investigate the relationship between MAPE leader and the experimental condition and a series of leaders' traits (openness, conscientiousness, and gender). As analyzed previously in the control variables section, openness and conscientiousness show stronger relationships with judgmental forecasting and cognitive ability in general in the literature. The regression in Table 5 explained 28.29% of the variance from the variable MAPE leader. An ANOVA was used to test whether this value significantly differed from zero and found statistical significance, $F=3.16$, $p = 0.025$, $R\text{-square} = 0.28$ (a moderate value in social studies).

The experimental condition variable exhibits a substantial negative effect on the outcome variable—smaller MAPE values represent higher accuracy—with a coefficient of $\beta_{\text{Exp_cond}} = -0.06$ ($t = -2.54$, $p = 0.016$). Thus, our empirical findings provide supporting evidence for hypothesis 1. On the other hand, the other predictor variables related to leader characteristics, namely Leader Big-5 (Cons), Leader Big-5 (Open), and Leader Gender Female, demonstrate negligible effects on the outcome variable.

Table 5: Regression results, p values<0.05

Model	Unstandardized Coefficients	Standardized Coefficients	Stand. error	t	p	95% confidence interval for B	
	B	Beta				lower bound	upper bound
(Constant)	0.22		0.07	3.01	0.005	0.07	0.37
Experimental condition Intervention	-0.06	-0.4	0.02	-2.54	0.016*	-0.11	-0.01
Leader Big-5 (Cons)	0.01	0.26	0.01	1.69	0.101	0	0.02
Leader Big-5 (Open)	0	0.05	0.01	0.34	0.736	-0.01	0.02
Leader Gender Female	-0.03	-0.2	0.02	-1.25	0.221	-0.08	0.02

5.3. Measuring team accuracy by aggregating participants' forecasts

We next test hypothesis 2. In particular, the data collected in step 1—the deliberation screen—allows answering questions about whether a simple aggregation of team members' judgmental forecasts will yield higher accuracy independently of the team leaders' final forecast submission. We defined MAPE average group and interval accuracy variables for that purpose. We use two variants of the interval metric Mean Scaled Interval Score (MSIS), one using all six forecasts (MSIS6) and one using only four (MSIS4), excluding the time series steep breaks (in forecast rounds 2 and 6). The latter tries to determine if the judgments' accuracy varies when the times series is more or less stable.

We ran t-tests to examine the statistical significance of the mean differences between control and intervention groups for three variables (see Table 6). MAPE Avg Particip passed both Kolmogorov-Smirnov and Shapiro-Wilk normality tests with p-values above 0.05. MSIS6 and MSIS4 passed the Kolmogorov-Smirnov test but did not pass the Shapiro-Wilk test; however, a closer inspection of the respective Q-Q plots showed that most of the points fall along the diagonal line with two outliers, which can explain the Shapiro-Wilk test rejection, which is more

sensitive to deviations (see Q-Q plots in the Appendix, section 4). Therefore, we conclude that we can carry on with the analysis. For MAPE Avg Particip, the t-test yielded a statistically significant difference in the means of the two groups with t-value= 2.23 (df = 37 and p = 0.032). For MSIS6, the test found a non-significant t-value of 0.88 (df = 38, p = 0.386), whereas, for MSIS4, the t-test revealed a significant t-value of 3.2 (df = 38, p = 0.003).

Table 6: T-tests for different aggregation metrics

		t	df	p (2-tailed)
MAPE Avg Particip	Equal variances	2.23	37	0.032
MSIS6	Equal variances	0.88	38	0.386
MSIS4	Equal variances	3.2	38	0.003

Additionally, we run three separate regressions, controlling for the leader's gender (Table 7). In this case, we also observe significant effects of the Experimental Condition for MAPE Avg Particip (moderate R-square = 0.18) and MSIS4 (moderate R-square = 0.29) results in the expected direction since negative standardized coefficients (betas) for the Experimental Condition mean a lower score, representing better accuracy. No significant effects were found for Gender. These results lead us to believe that the intervention likely impacted the accuracy of the aggregation metrics.

Table 7: Regression analyses for MAPE avg group, MSIS6 and MSIS4

Variable	Model	Unstand. Coeffic.	Standard. Coeffic.	Standard error	t	p	95% confidence interval for B	
		B	Beta				lower bound	upper bound
MAPE avg group	(Constant)	0.32		0.02	18.97	<.001	0.29	0.36
	Experimental condition Intervention	-0.05	-0.39	0.02	-2.53	0.016	-0.09	-0.01
	Gender Female	-0.03	-0.25	0.02	-1.62	0.114	-0.07	0.01
MSIS6	(Constant)	29.74		1.34	22.14	<.001	27.01	32.46
	Experimental condition Intervention	-2.13	-0.21	1.6	-1.33	0.192	-5.38	1.12
	Gender Female	-3.34	-0.31	1.65	-2.02	0.051	-6.69	0.01
MSIS4	(Constant)	42.57		4.18	10.18	<.001	34.09	51.05
	Experimental condition Intervention	-17.91	-0.5	4.99	-3.59	0.001	-28.02	-7.81
	Gender Female	-10.27	-0.28	5.15	-1.99	0.054	-20.71	0.17

5.4. Robustness checks

In order to determine the statistical validity of the intervention and the overall robustness of the previous findings, we use multivariate analysis of variance (MANOVA) including three groups of dependent variables: a) leader final forecast measured by MAPE leader (used for hypothesis one testing), together with two complementary metrics such as Mean Absolute Error (MAE) and Root Mean Square Error (RMSE), b) forecast aggregation metrics (used for hypothesis two testing) such as MAPE avg group and the two interval groups measurements MSIS4 and MSIS6; finally, c) variables measuring the level of percentual adjustment leaders exercise between step 1 (deliberation screen) and step 2 (leader screen). Leader_Self_Adj measures leaders' forecast change between the two steps, whereas Leader_Avg_Group measures the change between the mean forecast in Step 1 and the leader's final decision in Step 2. The assumption is that more mindful discussions in intervention groups will decrease the leaders' adjustments in Step 2.

All variables passed the Kolmogorov-Smirnov normality tests, whereas RSME_lead, MSIS6, and Leader Self and Avg adjustments did not pass the Shapiro-Wilk test. A closer inspection of the Q-Q plots revealed few outliers (see plots in the Appendix, section 4), and MANOVA is robust in addressing these violations (Ito, 1980), so we assumed the normality assumption is preserved. We then proceeded to analyze Box's Test of Equality of Covariance Matrices. Even though we obtained a significant value that would not pass the test, analyzing the log determinants values, we observed they are similar for the experimental conditions and the pooled within-groups (log determinants of -15.785, -20.292, and -15.330 for control, intervention, and pooled within groups respectively) allowing us to continue the analysis (Ito, 1980).

The one-way MANOVA revealed a significant multivariate main effect for the experimental condition (Exp_Cond), Pillai's Trace = 0.405, $F(8, 28) = 2.384$, $p < 0.05$, partial eta squared = 0.405 (see Table 8).

Table 8: Multivariate tests, experimental condition

	Effect	Value	F	Hypothesis df	Sig.	Partial Eta Squared
Intercept	Pillai's Trace	0.993	495.731	8.000	<0.001	0.993
Exp_Cond	Pillai's Trace	0.405	2.384	8.000	0.042	0.405

The power to detect the effect was 0.834; higher than 0.8 is considered high (Muller et al., 1992), where the number of groups (ng) in the experimental design is 2 (control and intervention), the number of dependent variables (nm) in the MANOVA test is 8 (MAPE_lead, MAE_lead, RSME_lead, MSIS_4_val, MAPE_lag, MAE_lag, RSME_lag, and Lead_Self_Adj_2to6), the significance level (alpha) is 0.05, the effect size (f) to detect with the test is 0.25 (medium effect size) and, the correlation (nscor) among the dependent variables is 0.7 (high correlation).

$$power = 1 - F_{1-\alpha, ng-1, (ng-1)(nm-1)} \left(\frac{(ng-1)(nm-1)}{ng(nm-1) + f^2(ng-1)^2} \right)$$

Where the effect size (f) depends on the nscor and where δ is the standardized mean difference between the groups:

$$f = \sqrt{\frac{nscor \times \delta^2}{1 - nscor}}$$

Given the significance of the overall MANOVA test, the univariate main effects of the experimental condition variable were examined (see Table 9). Significant univariate main effects for Exp_Cond were obtained for MAPE_lead, MAE_lead, MAPE_Avg_Particip, and Lead_Self_Adj_2to6. Significant pairwise differences between control and intervention groups were obtained with mean values of MAPE_lead (0.3100 vs. .2511), MAE_lead (73.7228 vs 63.5358), MAPE_Avg_Particip (0.3011 vs 0.2467), MSIS_4_val (32.8917 vs 22.2516), and Lead_Self_Adj_2to6 (.0550 vs .0253). We can conclude that the findings are robust, with a convergence in several accuracy metrics.

Table 9: Multivariate tests, dependent variables

Source	Dependent Variable	Type III Sum of Squares	F	Sig.	Partial Eta Squared
Exp_Cond	MAPE_lead	0.032	6.432	0.016	0.155
	MAE_lead	959.215	4.939	0.033	0.124
	RSME_lead	463.258	2.304	0.138	0.062
	MAPE_Avg_Particip	0.013	4.145	0.049	0.106
	MSIS_4_val	1046.441	7.119	0.011	0.169
	MSIS_6_val	0.372	0.027	0.871	0.001
	Lead_Self_Adj_2to6	0.008	4.199	0.048	0.107
	Lead_AvgGroup_Adj_2to6	0.003	1.375	0.249	0.038

Boxplot graphs for the accuracy variables showing significant effects are shown in Figure 6, depicting lower selected metrics scores and, hence, higher accuracy. Interestingly, all mean values (dotted lines) are above the median (solid lines), representing the impact of inaccurate teams in both control and intervention groups.

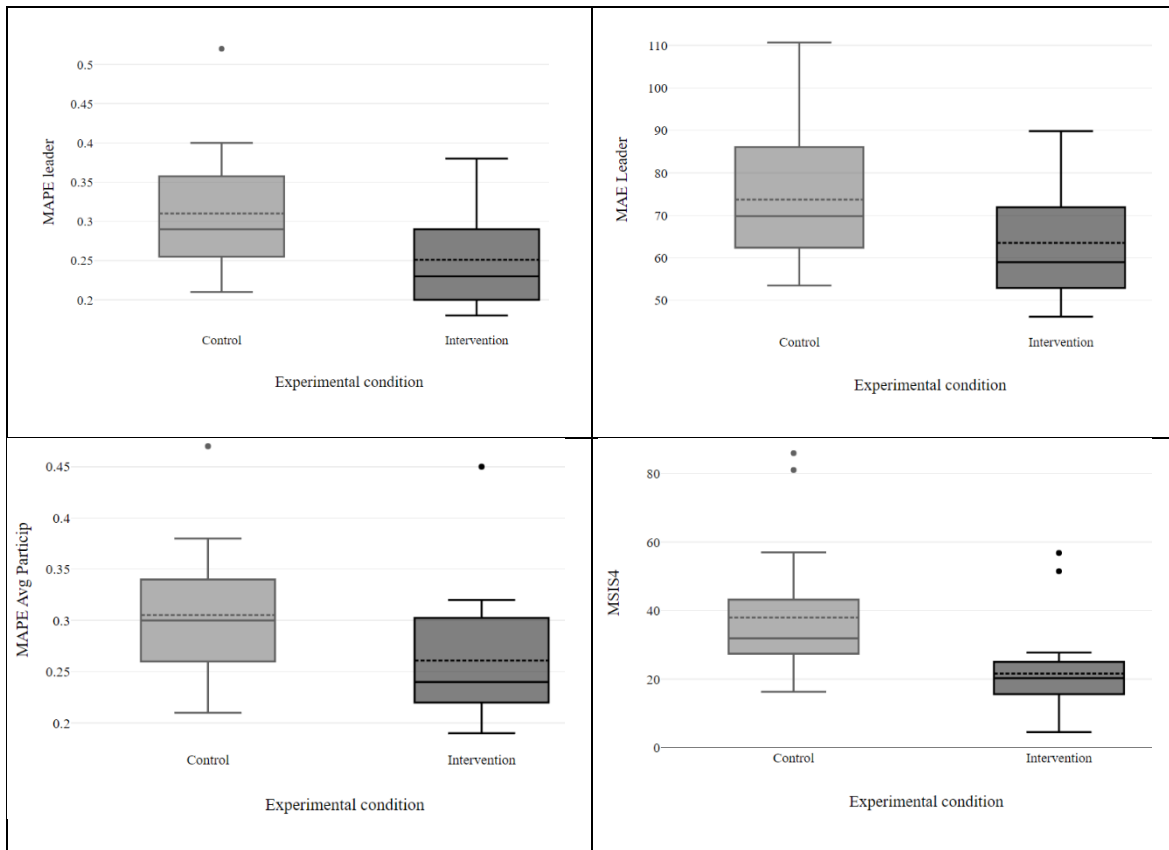


Figure 6: Pairwise mean difference and dispersion for accuracy metrics

5.5. Teams chat analysis

An additional analysis was conducted to shed light on the team communication dynamics in control and intervention groups using the dialogs between team members recorded in the chat functionality. The control groups registered an average of 259 words versus 326 in the intervention groups (29% higher). More specifically, the leaders in the control groups exhibited a word count of 94, versus the intervention group leaders, who exhibited a word count of 124 (33% higher). These metrics indicate more extended interactions between team members and could serve as a proxy of intervention leaders' facilitation efforts.

Regarding the content of these interactions, we chose to analyze the leader's use of language using Microsoft Word Copilot, a generative artificial intelligence feature that uses Open AI chat GPT 4.0. This tool helps interpret sentiment in written communication and yields language style analyses. In order to perform the sentiment analysis, all the leaders' chat communication in control and intervention groups were extracted and input to the algorithm separately in independent sessions to avoid cross-contamination.

Prompt: This text was extracted from a list of interactions between a leader and his team. Please a) analyze the overall sentiment in the text and b) analyze whether the leader facilitated the discussion, prompting opposing views, encouraging participation, and analyzing previous attempts.

Copilot answer (control group leaders): *Based on the analysis of the text from the document, the sentiment expressed seems to be a mix of uncertainty, agreement, and some negativity, as indicated by phrases like "I do not know what to think", "difficult" and "we are horrible. Haha". However, there are also instances of agreement, encouragement, and a willingness to collaborate, such as "agreed!" and "Go with your instinct team," which indicates positive sentiment.*

The leadership approach seems to be more facilitative, with phrases like "Go with your instinct team" and "I'm happy with 200 too" suggesting a collaborative and supportive environment. The leader also appears to have prompted the team to

consider different perspectives and analyze previous attempts, as indicated by comments such as "Lets concentrate on the black arrow" and "since we got both wrong," which suggest a reflective approach to decision-making.

Copilot answer (intervention group leaders): *The sentiment analysis of the text extracted from the interactions between the leader and their team indicates a positive and cooperative approach.*

The leader seems to have effectively encouraged participation, prompted for opposite views, and analyzed previous attempts, which aligns with the training they received in meeting facilitation. Phrases like "any different ideas?", "anyone have some suggestions?", "OPPO TO ME P2!!! OPPOSITE VIEW" and "Let's all submit our predictions now? Unless there is any more points from anyone?" demonstrate an inclusive and engaging leadership style. The leader also reflects on past experiences, as seen in phrases like "any previous mistakes we can learn from" and "what have we learnt so far from these sales thought?", which suggests a thoughtful and analytical approach to guiding the team's efforts. Overall, the text showcases a leader who is open to diverse perspectives and committed to collaborative problem-solving.

Even though the algorithm uses favorable terms to describe the leaders' role in both cases, it recognizes that leaders in the control groups have a certain degree of uncertainty and confusion, which would be expected, given that the control intervention was not designed to provide tools to manage short-term-live interactions. Conversely, it seems that intervention group leaders reflected the mindful organizing training objectives by using more nuanced language to direct the discussion and enabling discussion and participation more explicitly than control group leaders. This way of conducting the team facilitation coincides with the longer word count in the intervention groups for both the whole team and leader's word count.

6. DISCUSSION AND CONCLUSIONS

We aimed to examine the influence of team dynamics on collective forecasting performance. We adopted mindful organizing as a theoretical lens to understand how the interactions among forecasting team members could be structured. This perspective aligns with the literature identifying team interactions and teamwork—manifested through communication, coordination, and cooperation—as critical determinants of organizational effectiveness and performance (Kozlowski & Chao, 2012, 2018; Kozlowski & Ilgen, 2006). Consequently, our analysis emphasized team dynamics in contrast to similar studies in the field, which have taken a process-oriented view. In one exemplar of that approach, Oliva and Watson (2009, 2011) concluded that the planning process design determines collective behaviors such as information-sharing, validation, feedback, and learning among functional groups. Conversely, our approach deepens the level of analysis from the planning method to team micro-interactions, which we theorized plays a crucial role in avoiding superficial information processing. By engaging in a specific type of dialogue with others—directed towards the processes of mindful organizing, such as reluctance to oversimplify and attention to error—individuals can challenge their assumptions, clarify their goals, and explore alternative perspectives (Langer, 2000; Weick et al., 1999; Weick & Roberts, 1993) while avoiding the traps of overconfidence (Fischhoff et al., 1977).

Our experimental study showed that a mindful organizing intervention improved collective judgments. The regression analysis and robustness checks revealed that the experimental condition significantly improved accuracy by 24% vs. control groups (MAPE leader with $p = 0.018$). This result suggests that the intervention led to more accurate decision-making by the leaders. The study also found accuracy improvements in metrics of collective forecast aggregation—without the final leader intervention. We found significant differences between the two groups regarding MAPE Avg Particip and MSIS4 (excluding the time series peaks) but not MSIS6 (including all six observations). Additionally, team

communication dynamics analysis using Microsoft Word Copilot yielded a higher word count in intervention teams (29% higher), whereas the sentiment analysis algorithm identified instances of leader facilitation per the mindful organizing training.

From a scholarly perspective, our findings contribute to the literature on judgmental forecasting by infusing relevant collective decision-making theory in a field that has paid more attention to individual judgmental biases (Eroglu & Croxton, 2010; Flyvbjerg, 2008; Kahneman & Lovallo, 1993; Lawrence & O'Connor, 1995; Lee & Siemsen, 2016) groupthink and bandwagon effects (Janis, 1982; Mellers et al., 2014; Sniezek, 1989; Surowiecki, 2005) as well as and the potential bias accentuation effect of the group (Buehler et al., 2005). From a managerial perspective, our findings exemplify the relevance of leaders' roles in team outcomes, highlighting their function as combinatorial devices of the group's available information by acting as facilitators (Edmondson, 1996, 1999; Tannenbaum et al., 1998). The fact that personality and gender did not seem statistically significant in the results implies that these routines can be taught and implemented in knowledge work settings regardless of personal characteristics. Also, this study is interesting from a mindful organizing perspective since Sutcliffe et al. (2016) and Vogus & Sutcliffe (2012) have suggested that mindful organizing is a dynamic and fragile process that organizational actors must reconstitute continuously. In this context, they have implied that the time required for mindful organizing to emerge and persist will vary depending on the context, the level of interdependence, the organization's size, and the characteristics of the individuals involved. In this case, our study suggests that collective mindful states can emerge quickly in newly formed groups.

Our theoretical approach and experimental design are infrequent in the forecasting literature; hence, future studies must incorporate further validity provisions regarding replication and methodological refinement to overcome present limitations. Regarding sample characteristics, the study sample consisted of 120 participants grouped into 40 groups of three. While this sample size may be adequate for the conducted analyses, it might limit the

generalizability of the findings to larger populations or different group sizes. Future research could consider increasing the sample size and including participants from diverse backgrounds to enhance the external validity of the results. Also, this study used newly formed teams following Kozlowski & Chao (2012) observation that in previously formed teams, the most critical phenomena have already emerged; however, naturalistic organizations have a shared history that influences their interactions. Another aspect is using self-report measures to assess mindfulness and personality traits. Self-report measures are subject to various biases, including social desirability bias and response bias, which might influence the results' accuracy. Future studies could incorporate additional objective measures or alternative assessment methods to complement the self-report data. Finally, regarding outcome measures, the study focused primarily on the accuracy of team forecasts and the effects of the intervention on the MAPE leader variable. While this provides valuable insights into the specific research question, it may overlook other essential outcome variables related to team performance or the broader impact of mindfulness interventions. Future research could explore additional outcome variables, such as team decision-making quality, satisfaction, or long-term performance outcomes.

Future research

Further investigation could explore potential mediating or moderating factors influencing the relationship between the intervention and the accuracy of team forecasts. For example, factors like team cohesion, communication patterns, or individual characteristics of team members may interact with the intervention's effectiveness. Understanding these underlying mechanisms could provide a more comprehensive understanding of how mindfulness interventions impact team performance. Additionally, the presented data primarily focuses on the immediate effects of the intervention on team accuracy. However, examining the sustainability of the intervention's impact over an extended period could offer valuable insights, especially since there is no consensus in the literature about how long it takes mindful organizing to emerge and whether it can be sustained (Sutcliffe et al., 2016). Longitudinal studies or follow-up assessments could help

determine whether the observed improvements in team accuracy persist over time or diminish after the intervention period. Finally, exploring the boundary conditions of the intervention's effectiveness can further enhance the understanding of its applicability. Factors such as team size, task complexity, or organizational context might influence the outcomes of the intervention. Future research could investigate these boundary conditions to identify when and for whom mindfulness interventions are most effective in enhancing team accuracy.

In conclusion, studying the impact of organization-level characteristics on forecasting performance remains an important topic likely to yield theory and practice-relevant insights. Specifically, these results bring a new direction in the conversation about collective judgmental performance in forecasting tasks, where focusing on the quality of the team's interactions and its manipulation become the primary consideration. This insight invites us to reshuffle attention toward the forecasting process as an organizational phenomenon where accuracy is a byproduct of adaptive team processes (Kozlowski & Chao, 2012; Kozlowski & Ilgen, 2006). Also, this approach reinstates the observation that superficial information search and processing are the primary source of bias in individual and collective judgment (Fischhoff et al., 1977; Slovic & Lichtenstein, 1971), which can be counteracted with heedful and dense interrelations (Weick et al., 1999; Weick & Roberts, 1993).

IV. DISCUSSION AND CONCLUSIONS

The research question explored in this thesis is how to improve collective forecast accuracy by enhancing team interactions. The empirical study focused on mindful organizing principles and their impact on team forecasting accuracy. The study found that interventions based on mindful organizing significantly improved team leader forecasts by 24%, measured by MAPE leader (other point accuracy metrics such as MAPE avg. participants also showed significant results). However, interval accuracy scores did not show a significant relationship.

This work started with insight into the lack of theoretical frames to understand forecasting from an organizational perspective. This lead was followed by exploring the extant literature, focusing on forecast reviews of the last 40 years, forecasting studies addressing organizational elements, literature on organizational learning, and literature on high reliability, sensemaking, and mindful organizing. From that process, the study identified the limiting effects of the extant paradigm—functionalism—in some forecasting research streams situated in the adjacency of organizations and teams, such as extracting forecasts from groups and learning from forecast feedback. Additionally, this process drove the conclusion that potential frameworks must introduce organizational nuances such as the hierarchical nature of teams, the importance of the leader's role, performance deliberation through after-the-fact rituals, and the importance of disclosing mistakes. Finally, borrowing from the literature on organizational learning and high-reliability organizations, mindful organizing has been put forward as a potential adaptive theoretical frame for collective forecasting tasks, given its focus on the quality of organizational attention, communication micro-processes, and a focus on deliberation.

The empirical research process involved an experimental study with hierarchical triads. The study aimed to highlight two aspects of real-life teams and situations: the existence of a hierarchical structure and the need to intervene at the level of team communication, making team interactions the primary level of intervention. Hence, the mindful organizing intervention emphasized team communication and

facilitation strategies. Participants were randomly assigned to either the intervention or control conditions to forecast six points of an artificially created time series, forming 40 hierarchical triads. In the intervention group, participants received training emphasizing mindful organizing strategies, particularly team communication and facilitation. Meanwhile, the control groups watched decoy videos. By comparing the outcomes between these groups, the study assessed the effectiveness of mindful organizing in enhancing team leaders and whole team forecast accuracy metrics.

From a scholarly perspective, our findings contribute to the literature on judgmental forecasting by infusing relevant collective decision-making theory in a field that has paid more attention to individual judgmental biases (Eroglu & Croxton, 2010; Flyvbjerg, 2008; Kahneman & Lovallo, 1993; Lawrence & O'Connor, 1995; Lee & Siemsen, 2016) groupthink and bandwagon effects (Janis, 1982; Mellers et al., 2014; Sniezek, 1989; Surowiecki, 2005) as well as and the potential bias accentuation effect of the group (Buehler et al., 2005). Also, it is vital to acknowledge that team interactions are the currency of team cognitive performance, avoiding falling into the trap of a dichotomy of the adaptive nature of human judgment and making it contingent on the quality and structure of team micro-processes and interactions.

From a managerial perspective, our findings exemplify the relevance of leaders' roles in team outcomes, highlighting their function as combinatorial devices of the group's available information by acting as facilitators (Edmondson, 1996, 1999; Tannenbaum et al., 1998). Most teams in natural settings operate with leader-subordinate relationships, and, to our knowledge, no studies in forecasting have specifically addressed these dynamics. Here, the leader as facilitator insight provides an actionable framework for improving team performance with simple training and upskilling, borrowing from the currently widespread literature on psychological safety (Edmondson, 1996). The fact that personality and gender did not seem statistically significant in the results implies that these routines can be taught and implemented in knowledge work settings regardless of personal characteristics.

Also, this study is interesting from a mindful organizing perspective since Sutcliffe et al. (2016) and Vogus & Sutcliffe (2012) have suggested that mindful organizing is a dynamic and fragile process that organizational actors must reconstitute continuously. In this context, they have implied that the time required for mindful organizing to emerge and persist will vary depending on the context, the level of interdependence, the organization's size, and the characteristics of the individuals involved. In this case, our study suggests that collective mindful states can emerge quickly in newly formed groups.

In conclusion, mindful organizing principles can enhance team forecasting accuracy. Organizations should foster learning cultures, encourage error disclosure, and recognize the value of successes and failures.

(traducción al español)

La pregunta de investigación explorada en esta tesis es cómo mejorar la precisión de las previsiones colectivas mediante la mejora de las interacciones del equipo. El estudio empírico se centró en los principios de organización consciente y su impacto en la precisión de las previsiones del equipo. El estudio encontró que las intervenciones basadas en la organización consciente mejoraron significativamente las previsiones del líder del equipo en un 24%, medido por el MAPE del líder (otros indicadores de precisión, como el MAPE promedio de los participantes, también mostraron resultados significativos). Sin embargo, los puntajes de precisión de intervalo no mostraron una relación significativa.

Este trabajo comenzó con una comprensión de la falta de marcos teóricos para comprender las previsiones desde una perspectiva organizativa. Esta pista fue seguida por la exploración de la literatura existente, centrándose en las revisiones de previsiones de los últimos 40 años, estudios de previsión que abordan elementos organizativos, literatura sobre aprendizaje organizativo y literatura sobre alta confiabilidad, construcción de sentido y organización consciente. A partir de ese proceso, el estudio identificó los efectos limitantes del paradigma existente, el funcionalismo, en algunas corrientes de investigación de previsión situadas en la adyacencia de organizaciones y equipos, como la extracción de previsiones de grupos y el aprendizaje a partir de la retroalimentación de las previsiones. Además, este proceso llevó a la conclusión de que los marcos potenciales deben introducir matices organizativos, como la naturaleza jerárquica de los equipos, la importancia del papel del líder, la deliberación del rendimiento a través de rituales posteriores al hecho y la importancia de divulgar errores. Finalmente, tomando prestado de la literatura sobre aprendizaje organizativo y organizaciones de alta confiabilidad, la organización consciente se ha propuesto como un marco teórico adaptativo potencial para las tareas de previsión colectiva, dada su atención a la calidad de

la atención organizativa, los microprocesos de comunicación y el enfoque en la deliberación.

El proceso de investigación empírica involucró un estudio experimental con tríadas jerárquicas. El estudio tenía como objetivo resaltar dos aspectos de los equipos y situaciones de la vida real: la existencia de una estructura jerárquica y la necesidad de intervenir a nivel de la comunicación del equipo, convirtiendo las interacciones del equipo en el nivel principal de intervención. Por lo tanto, la intervención de organización consciente enfatizó las estrategias de comunicación y facilitación del equipo. Utilizando un diseño de 40 tríadas jerárquicas, los participantes fueron asignados al azar a las condiciones de intervención o control para predecir seis puntos de una serie temporal creada artificialmente. En el grupo de intervención, los participantes recibieron capacitación enfocada en estrategias de organización consciente, especialmente en comunicación y facilitación del equipo. Mientras tanto, los grupos de control vieron videos de distracción. Al comparar los resultados entre estos grupos, el estudio evaluó la efectividad de la organización consciente en mejorar las métricas de precisión de las previsiones del líder del equipo y del equipo completo.

Desde una perspectiva académica, nuestros hallazgos contribuyen a la literatura sobre pronósticos basados en juicios al infundir teoría relevante de toma de decisiones colectivas en un campo que ha prestado más atención a los sesgos de juicio individuales (Eroglu & Croxton, 2010; Flyvbjerg, 2008; Kahneman y Lovallo, 1993; Lawrence y O'Connor, 1995), el pensamiento grupal y los efectos de la moda (Janis, 1982; Mellers et al., 2014; Sniezek, 1989; Surowiecki, 2005), así como el efecto potencial de acentuación de sesgos del grupo (Buehler et al., 2005). Además, es vital reconocer que las interacciones del equipo son la moneda del rendimiento cognitivo del equipo, evitando caer en la trampa de una dicotomía de la naturaleza adaptativa del juicio humano y haciéndola dependiente de la calidad y estructura de los microprocesos e interacciones del equipo.

Desde una perspectiva gerencial, nuestros hallazgos ejemplifican la relevancia de los roles de los líderes en los resultados del equipo, resaltando su función como dispositivos combinatorios de la información disponible del grupo al actuar como facilitadores (Edmondson, 1996, 1999; Tannenbaum et al., 1998). La mayoría de los equipos en entornos naturales operan con relaciones líder-subordinado, y, según nuestro conocimiento, no se han abordado específicamente estas dinámicas en los estudios de pronóstico. Aquí, la idea del líder como facilitador proporciona un marco de acción para mejorar el rendimiento del equipo con capacitación y desarrollo relativamente sencillos, tomando prestado de la amplia literatura sobre seguridad psicológica (Edmondson, 1996). El hecho de que la personalidad y el género no parecieran estadísticamente significativos en los resultados implica que estas rutinas pueden enseñarse e implementarse en entornos de trabajo del conocimiento independientemente de las características personales.

Además, este estudio es interesante desde una perspectiva de organización consciente, ya que Sutcliffe et al. (2016) y Vogus & Sutcliffe (2012) han sugerido que la organización consciente es un proceso dinámico y frágil que los actores organizativos deben reconstituir continuamente. En este contexto, han insinuado que el tiempo necesario para que la organización consciente emerja y persista variará según el contexto, el nivel de interdependencia, el tamaño de la organización y las características de las personas involucradas. En este caso, nuestro estudio sugiere que los estados conscientes colectivos pueden surgir rápidamente en grupos recién formados.

En conclusión, los principios de organización consciente pueden mejorar la precisión de las previsiones del equipo. Las organizaciones deben fomentar culturas de aprendizaje, alentar la divulgación de errores y reconocer el valor de los éxitos y fracasos.

V. APPENDIX

1. Instructions

Time left to complete this page: 4:40

Instructions

You are part of a team that has been tasked with producing 6 forecasts for ACME Inc's monthly sales.

Team configuration

- Each team will have 3 members: 1 team leader assigned randomly and 2 team participants. These roles will remain constant until the end of the task.
- All the team members must discuss their opinions about the forecast in the team chat.
- Once the discussion—or the time limit—is over, all team members will submit an individual forecast.
- **Importantly, only the leader's final forecast will count for the group scoring.** The leader will have the opportunity to make a final decision in a separate screen that will show all individual forecasts.

The task (estimated duration: 45mins)

- Before starting the forecasting task, you must watch a short video(s) and answer some simple comprehension questions.
- Your team needs to provide 6 forecasts for monthly sales in 6 rounds.
- The process is as follows:
 - Round 1: You get 48 months of actuals as input and must forecast the 49th month.
 - Round 2: You get the 49 months of actuals (including the comparison between your round 1 forecast and the actuals) and must forecast the 50th month.

Measuring performance

- After each round you will see the group forecast (leader's decision) compared to the actuals so you can discuss potential adjustments in the next round.
- Once the 6 rounds are completed, your team will get an accuracy score using MAPE (mean absolute percentage error):
 - less than 20% MAPE value means a good or highly accurate forecast,
 - between 20-50% reasonable,
 - and more than 50% is inaccurate.
- Example of MAPE scoring (recall: the lower the better):

	Forecast	Actual	Deviation	Absolute Deviation	% of error
Round 1	125	120	-5	5	4%
Round 2	125	130	5	5	4%
Round 3	125	110	-15	15	14%
Round 4	125	140	15	15	11%
Round 5	125	110	-15	15	14%
Round 6	125	130	5	5	4%
TOTAL					49.8%
MAPE					8.3%

Payoff

The full participation will grant each participant the maximum reward of £5.00 dollar for each participant, to this amount we will deduct the MAPE score. In the previous example, an individual payout will be calculated as $£5.00 \times (100\% - 8.3\%) = £4.59$

Next

2. Intervention videos

Video 1 out of 4: Watch this video before to proceed to the next page

Time left to complete this page: 0:31

Attention: The 'Next' button will be visible as soon as the video ends.

What is the role of the leader?

• Supervisor



• Facilitator



Answer the questions regarding the video you have just seen

Time left to complete this page: 2:48

What role could a team leader play to improve team task results?


- Supervisor
- Bold leader
- Visionary leader
- Facilitator

Next


intervention prolific beta chapter 1 Copy link


What is the role of the leader?

• Supervisor



• Facilitator



Watch on  YouTube

Video 2 out of 4: Watch this video before to proceed to the next page

Time left to complete this page: 0:47

Attention: The 'Next' button will be visible as soon as the video ends.



Answer the questions regarding the video you have just seen

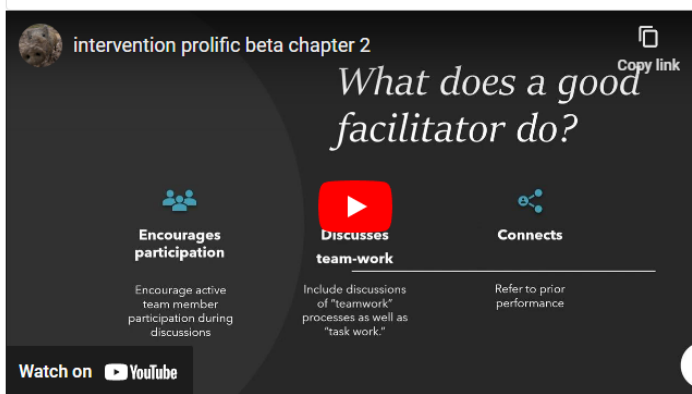
Time left to complete this page: 2:57

What constitutes effective facilitation strategies?

(select TWO choices)

- Ensure participation
- Assigning tasks effectively.
- Connect current task to previous performance.
- Being assertive and clear in providing direction.

Next



Video 3 out of 4: Watch this video before to proceed to the next page

Time left to complete this page: **0:50**

Attention: The 'Next' button will be visible as soon as the video ends.

Interacting like aircraft carriers' crews

- What did we assume that didn't happen?
- What are we missing?
- Are we fully immersed in the task?
- What actions can be adjusted and tested in the next round?
- Is someone in the team better at this task and should we let this person decide?



Answer the questions regarding the video you have just seen

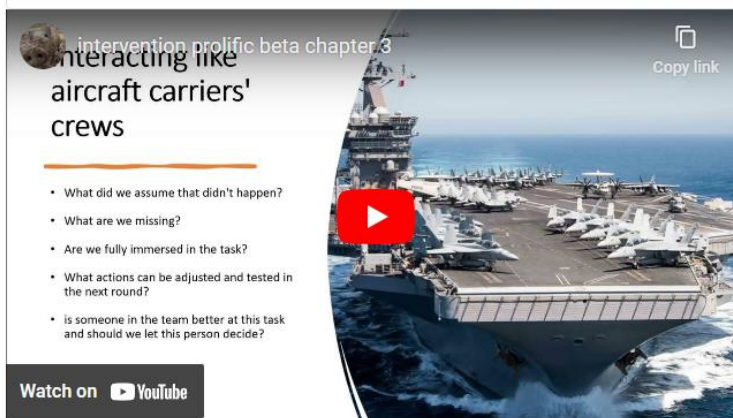
Time left to complete this page: **2:50**

How aircraft carriers' crews conduct discussions?

(select **TWO** choices)

- They focus on their successes and reassure each other.
- They focus on error and what could go wrong.
- They are confident in their leader's decisions and follow suit.
- They could defer decisions to team members.

Next



Video 4 out of 4: Watch this video before to proceed to the next page

Time left to complete this page: 0:26

Attention: The 'Next' button will be visible as soon as the video ends.

In summary



Facilitate instead of supervise

- If you are the team leader, act as a facilitator to get better results.



Follow facilitation tips

- Ensure participation, discuss team-work, and use previous performance as input to current task.



Discuss like an aircraft carrier crew

- When thinking about the content of the discussions, focus on error, blind spots, and defer some decisions to an expert team member.

Video 4 out of 4: Watch this video before to proceed to the next page

Time left to complete this page: 0:26

Attention: The 'Next' button will be visible as soon as the video ends.

In summary



Facilitate instead of supervise

- If you are the team leader, act as a facilitator to get better results.



Follow facilitation tips

- Ensure participation, discuss team-work, and use previous performance as input to current task.



Discuss like an aircraft carrier crew

- When thinking about the content of the discussions, focus on error, blind spots, and defer some decisions to an expert team member.

3. Forecasting game interfaces

Full instructions

Short instructions

- All team members must submit a forecast, but the leader is responsible for deciding the final team forecast and submitting it (in a separate screen).
- Your team needs to provide 6 forecasts for monthly sales, one month each time during 6 rounds.
- The next round starts only when all 3 members have submitted, and the leader has made the final decision.
- Once the 6 rounds are completed, your team will get an accuracy score using MAPE (mean absolute percentage error):
 - less than 20% MAPE value means a good or highly accurate forecast,
 - between 20-50% reasonable,
 - and more than 50% is inaccurate.

Sales of ACME Corp.

Previous values

Please chat with other group members first in the box below:

Leader good morning
Participant 1 hello guys
Participant 2 (Me) What do you think?

Insert your personal prediction in the box below!

Info:

Time left to complete this page: 9:10

Round number:

Your role:

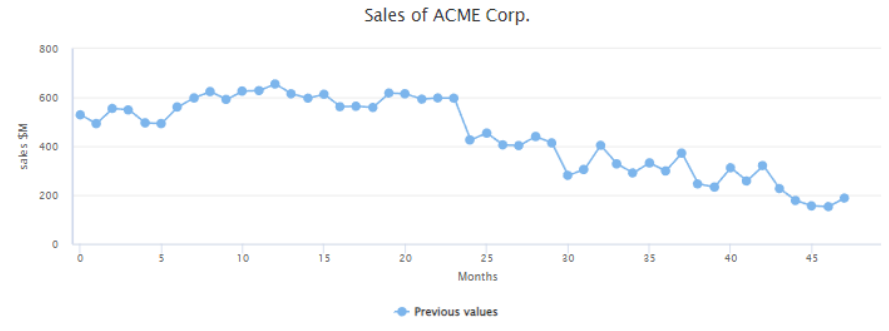
Your prediction:

(At the box above you can submit as many predictions as you want)

Leader decision

Time left to complete this page: 1:50

Remember, good leaders facilitate discussions, learn from previous mistakes, and defer decisions to expert team members.



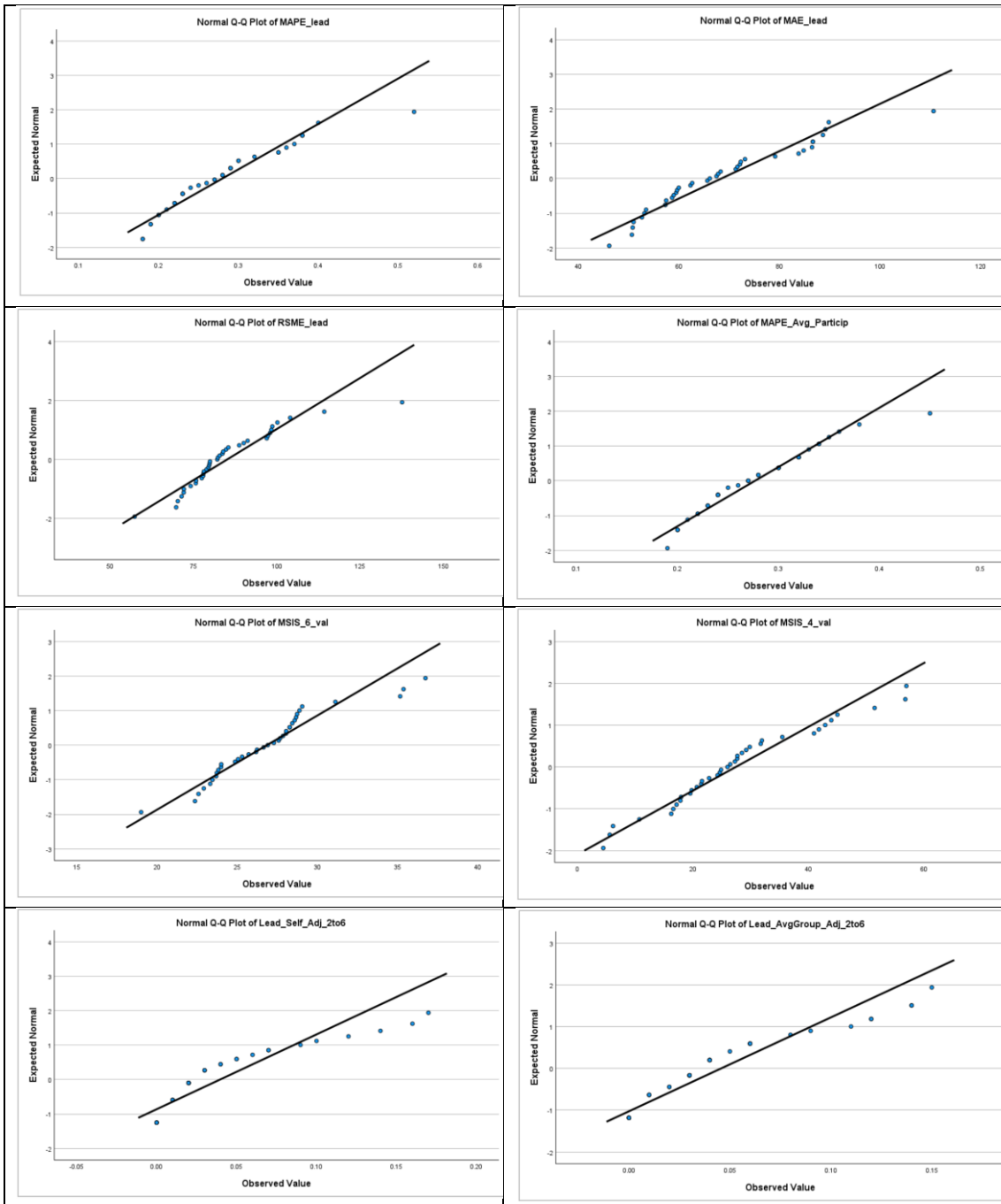
Make a final prediction

Participant:	Prediction:
Leader (you)	200
Participant 1	220
Participant 2	200

Look at the predictions made by you at the discussion stage and make the final prediction for this round:

Next

4. Q-Q plots for normality tests



VI. REFERENCES

- 2015 *Breakthrough Innovation Report, Europe Edition*. (n.d.). Retrieved May 31, 2020, from <http://innovation.nielsen.com/breakthrough2015EU>
- Alavi, M., & Leidner, D. (1999). Knowledge Management Systems: Issues, Challenges, and Benefits. *Communications of the Association for Information Systems*, 1(1). <https://doi.org/10.17705/1CAIS.00107>
- Alavi, M., & Leidner, D. E. (2001). Review: Knowledge Management and Knowledge Management Systems: Conceptual Foundations and Research Issues. *MIS Quarterly*, 25(1), 107–136. JSTOR. <https://doi.org/10.2307/3250961>
- Argote, L. (1999). *Organizational Learning: Creating, Retaining and Transferring Knowledge*. Springer Science & Business Media.
- Argote, L., & Miron-Spektor, E. (2011). Organizational Learning: From Experience to Knowledge. *Organization Science*, 22(5), 1123–1137. <https://doi.org/10.1287/orsc.1100.0621>
- Armstrong, J.S. (1988). Research needs in forecasting. *International Journal of Forecasting*, 4(3), 449–465. [https://doi.org/10.1016/0169-2070\(88\)90111-2](https://doi.org/10.1016/0169-2070(88)90111-2)
- Athanasopoulos, G., & Hyndman, R. J. (2011). The value of feedback in forecasting competitions. *International Journal of Forecasting*, 27(3), 845–849. <https://doi.org/10.1016/j.ijforecast.2011.03.002>
- Ausserhofer, D., Schubert, M., Desmedt, M., Blegen, M. A., De Geest, S., & Schwendimann, R. (2013). The association of patient safety climate and nurse-related organizational factors with selected patient outcomes: A cross-sectional survey. *International Journal of Nursing Studies*, 50(2), 240–252. <https://doi.org/10.1016/j.ijnurstu.2012.04.007>

- Barney, J. (1991). Firm Resources and Sustained Competitive Advantage. *Journal of Management*, 17(1), 99–120. <https://doi.org/10.1177/014920639101700108>
- Barry, D., & Meisiek, S. (2010). Seeing More and Seeing Differently: Sensemaking, Mindfulness, and the Workarts. *Organization Studies*, 31(11), 1505–1530. <https://doi.org/10.1177/0170840610380802>
- Bass, B. M., & Avolio, B. J. (1993). Transformational Leadership and Organizational Culture. *Public Administration Quarterly*, 17(1), 112–121.
- Bass, B. M., & Riggio, R. E. (2005). *Transformational Leadership* (2nd ed.). Psychology Press. <https://doi.org/10.4324/9781410617095>
- Baumard, P., & Starbuck, W. H. (2005). Learning from failures: Why it May Not Happen. *Long Range Planning*, 38(3), 281–298. <https://doi.org/10.1016/j.lrp.2005.03.004>
- Bedeian, A. G. (2002). The Dean's Disease: How the Darker Side of Power Manifests Itself in the Office of Dean. *Academy of Management Learning & Education*, 1(2), 164–173. <https://doi.org/10.5465/amle.2002.8509359>
- Bloch, M., Blumberg, S., & Laartz, J. (2012). *Delivering large-scale IT projects on time, on budget, and on value*. 27, 6.
- Bolger, F., & Önkal-Atay, D. (2004). The effects of feedback on judgmental interval predictions. *International Journal of Forecasting*, 20(1), 29–39. [https://doi.org/10.1016/S0169-2070\(03\)00009-8](https://doi.org/10.1016/S0169-2070(03)00009-8)
- Bolger, F., & Wright, G. (1994). Assessing the quality of expert judgment: Issues and analysis. *Decision Support Systems*, 11(1), 1–24. [https://doi.org/10.1016/0167-9236\(94\)90061-2](https://doi.org/10.1016/0167-9236(94)90061-2)
- Bossche, P. van den, Gijssels, W., Segers, M., Woltjer, G. B., & Kirschner, P. (2011). Team learning: Building shared mental models. *Instructional Science*, 39(3), 283–301. <https://doi.org/10.1007/s11251-010-9128-3>

- Bretschneider, S., Gorr, W. L., Grizzle, G., & Klay, E. (1989). Political and organizational influences on the accuracy of forecasting state government revenues. *International Journal of Forecasting*, 5(3), 307–319.
- Browning, L., & Boudes, T. (2005). The use of narrative to understand and respond to complexity: A comparative analysis of the Cynefin and Weickian models. *Emergence: Complexity and Organization*, 7(3–4), 35–43.
- Buehler, R., Messervey, D., & Griffin, D. (2005). Collaborative planning and prediction: Does group discussion affect optimistic biases in time estimation? *Organizational Behavior and Human Decision Processes*, 97(1), 47–63.
<https://doi.org/10.1016/j.obhdp.2005.02.004>
- Buhrmester, M., Kwang, T., & Gosling, S. D. (2016). *Amazon's Mechanical Turk: A new source of inexpensive, yet high-quality data?* (p. 139). American Psychological Association.
<https://doi.org/10.1037/14805-009>
- Burns, J. M. (1978). *Leadership*. Harper & Row.
- Burrell, G., & Morgan, G. (1979). *Sociological paradigms and organisational analysis: Elements of the sociology of corporate life*. Heinemann.
- Cannon, M. D., & Edmondson, A. C. (2005). Failing to Learn and Learning to Fail (Intelligently): How Great Organizations Put Failure to Work to Innovate and Improve. *Long Range Planning*, 38(3), 299–319. <https://doi.org/10.1016/j.lrp.2005.04.005>
- Cannon-Bowers, J. A., & Salas, E. (1998). Team performance and training in complex environments: Recent findings from applied research. *Current Directions in Psychological Science*, 7(3), 83–87. <https://doi.org/10.1111/1467-8721.ep10773005>

- Christianson, M. K., Farkas, M. T., Sutcliffe, K. M., & Weick, K. E. (2009). Learning through rare events: Significant interruptions at the Baltimore & Ohio Railroad Museum. *Organization Science*, *20*(5), 846–860. <https://doi.org/10.1287/orsc.1080.0389>
- Cokely, E. T., Feltz, A., Ghazal, S., Allan, J. N., Petrova, D., & Garcia-Retamero, R. (2018). Skilled decision theory: From intelligence to numeracy and expertise. In *The Cambridge handbook of expertise and expert performance, 2nd ed* (pp. 476–505). Cambridge University Press. <https://doi.org/10.1017/9781316480748.026>
- Cokely, E. T., & Kelley, C. M. (2009). Cognitive abilities and superior decision making under risk: A protocol analysis and process model evaluation. *Judgment and Decision Making*, *4*(1), 20–33.
- Collyer, S. C., & Malecki, G. S. (1998). Tactical decision making under stress: History and overview. In *Making decisions under stress: Implications for individual and team training* (pp. 3–15). American Psychological Association. <https://doi.org/10.1037/10278-016>
- Craske, M. G., Treanor, M., Conway, C. C., Zbozinek, T., & Vervliet, B. (2014). Maximizing exposure therapy: An inhibitory learning approach. *Behaviour Research and Therapy*, *58*, 10–23. <https://doi.org/10.1016/j.brat.2014.04.006>
- Crump, M. J. C., McDonnell, J. V., & Gureckis, T. M. (2013). Evaluating Amazon’s Mechanical Turk as a Tool for Experimental Behavioral Research. *PLOS ONE*, *8*(3), e57410. <https://doi.org/10.1371/journal.pone.0057410>
- Davenport, T. H., Davenport, P. D. of M. I. S. of M. T. H., & Prusak, L. (1998). *Working Knowledge: How Organizations Manage what They Know*. Harvard Business Press.
- Davies, S. (2006). Aesthetic Judgements, Artworks and Functional Beauty. *The Philosophical Quarterly*, *56*(223), 224–241. <https://doi.org/10.1111/j.1467-9213.2006.00439.x>

- Davis, D. F., & Mentzer, J. T. (2007). Organizational factors in sales forecasting management. *International Journal of Forecasting*, 23(3), 475–495.
- Desai, V. (2014). Learning Through the Distribution of Failures within an Organization: Evidence from Heart Bypass Surgery Performance. *Academy of Management Journal*, 58(4), 1032–1050. <https://doi.org/10.5465/amj.2013.0949>
- Deschamps, E. (2004). The impact of institutional change on forecast accuracy: A case study of budget forecasting in Washington State. *International Journal of Forecasting*, 20(4), 647–657. <https://doi.org/10.1016/j.ijforecast.2003.11.009>
- Dunbar, R. L. M., & Garud, R. (2009). Distributed Knowledge and Indeterminate Meaning: The Case of the Columbia Shuttle Flight. *Organization Studies*, 30(4), 397–421. <https://doi.org/10.1177/0170840608101142>
- Edmondson. (1996). Learning from Mistakes is Easier Said Than Done: Group and Organizational Influences on the Detection and Correction of Human Error. *The Journal of Applied Behavioral Science*, 32(1), 5–28. <https://doi.org/10.1177/0021886396321001>
- Edmondson. (1999). Psychological Safety and Learning Behavior in Work Teams. *Administrative Science Quarterly*, 44(2), 350–383. <https://doi.org/10.2307/2666999>
- Eggers, J. P., & Song, L. (2014). Dealing with Failure: Serial Entrepreneurs and the Costs of Changing Industries Between Ventures. *Academy of Management Journal*, 58(6), 1785–1803. <https://doi.org/10.5465/amj.2014.0050>
- Einhorn, H. J., & Hogarth, R. M. (1981). Behavioral Decision Theory: Processes of Judgment and Choice. *Journal of Accounting Research*, 19(1), 1–31. JSTOR. <https://doi.org/10.2307/2490959>

- Epley, N., & Gilovich, T. (2006). The Anchoring-and-Adjustment Heuristic: Why the Adjustments Are Insufficient. *Psychological Science*, *17*(4), 311–318. <https://doi.org/10.1111/j.1467-9280.2006.01704.x>
- Eroglu, C., & Croxton, K. L. (2010). Biases in judgmental adjustments of statistical forecasts: The role of individual differences. *International Journal of Forecasting*, *26*(1), 116–133. <https://doi.org/10.1016/j.ijforecast.2009.02.005>
- Fahimnia, B., Pournader, M., Siemsen, E., Bendoly, E., & Wang, C. (2019). Behavioral Operations and Supply Chain Management—A Review and Literature Mapping. *Decision Sciences*, *50*(6), 1127–1183. <https://doi.org/10.1111/dec.12369>
- Fildes, R. (2006). The forecasting journals and their contribution to forecasting research: Citation analysis and expert opinion. *International Journal of Forecasting*, *3*(22), 415–432. <https://doi.org/10.1016/j.ijforecast.2006.03.002>
- Fildes, R., Bretschneider, S., Collopy, F., Lawrence, M., Stewart, D., Winklhofer, H., Mentzer, J. T., & Moon, M. A. (2003). Researching sales forecasting practice commentaries and authors' response on "conducting a sales forecasting audit" by M.A. Moon, J.T. Mentzer & C.D. Smith. *International Journal of Forecasting*, *19*(1), 27–42. [https://doi.org/10.1016/S0169-2070\(02\)00033-X](https://doi.org/10.1016/S0169-2070(02)00033-X)
- Fildes, R., Goodwin, P., Lawrence, M., & Nikolopoulos, K. (2009). Effective forecasting and judgmental adjustments: An empirical evaluation and strategies for improvement in supply-chain planning. *International Journal of Forecasting*, *25*(1), 3–23. <https://doi.org/10.1016/j.ijforecast.2008.11.010>
- Fildes, R., & Hastings, R. (1994). The Organization and Improvement of Market Forecasting. *Journal of the Operational Research Society*, *45*(1), 1–16. <https://doi.org/10.1057/jors.1994.1>

- Finkelstein, S. (2004). *Why Smart Executives Fail: And What You Can Learn from Their Mistakes*. Penguin.
- Fiol, C. M., & O'Connor, E. J. (2003). Waking Up! Mindfulness in the Face of Bandwagons. *Academy of Management Review*, 28(1), 54–70. <https://doi.org/10.5465/amr.2003.8925227>
- Fischhoff, B., Slovic, P., & Lichtenstein, S. (1977). Knowing with certainty: The appropriateness of extreme confidence. *Journal of Experimental Psychology: Human Perception and Performance*, 3(4), 552–564. <https://doi.org/10.1037/0096-1523.3.4.552>
- Flyvbjerg, B. (2008). Curbing Optimism Bias and Strategic Misrepresentation in Planning: Reference Class Forecasting in Practice. *European Planning Studies*, 16(1), 3–21. <https://doi.org/10.1080/09654310701747936>
- Flyvbjerg, B. (2014). What you Should Know about Megaprojects and Why: An Overview. *Project Management Journal*, 45(2), 6–19. <https://doi.org/10.1002/pmj.21409>
- Gaba, D. M., Howard, S. K., Fish, K. J., Smith, B. E., & Sowb, Y. A. (2001). Simulation-Based Training in Anesthesia Crisis Resource Management (ACRM): A Decade of Experience: *Simulation & Gaming*. <https://doi.org/10.1177/104687810103200206>
- Galbraith, C. S., & Merrill, G. B. (1996). The Politics of Forecasting: Managing the Truth. *California Management Review*, 38(2), 29–43. <https://doi.org/10.2307/41165831>
- Galbraith, J. R. (1974). Organization Design: An Information Processing View. *INFORMS Journal on Applied Analytics*, 4(3), 28–36. <https://doi.org/10.1287/inte.4.3.28>
- GAO-15-675T, U. S. G. A. (2015). *Information Technology: Additional Actions and Oversight Urgently Needed to Reduce Waste and Improve Performance in Acquisitions and Operations*. GAO-15-675T. <https://www.gao.gov/products/gao-15-675t>

- Garvin, D. A., Edmondson, A. C., & Gino, F. (2008, March 1). Is Yours a Learning Organization? *Harvard Business Review*, March 2008. <https://hbr.org/2008/03/is-yours-a-learning-organization>
- Gavetti, G., Levinthal, D. A., & Rivkin, J. W. (2005). Strategy making in novel and complex worlds: The power of analogy. *Strategic Management Journal*, 26(8), 691–712. <https://doi.org/10.1002/smj.475>
- Gick, M. L., & Holyoak, K. J. (1980). Analogical problem solving. *Cognitive Psychology*, 12(3), 306–355. [https://doi.org/10.1016/0010-0285\(80\)90013-4](https://doi.org/10.1016/0010-0285(80)90013-4)
- Gick, M. L., & Holyoak, K. J. (1983). Schema induction and analogical transfer. *Cognitive Psychology*, 15(1), 1–38. [https://doi.org/10.1016/0010-0285\(83\)90002-6](https://doi.org/10.1016/0010-0285(83)90002-6)
- Gioia, D. A., & Pitre, E. (1990). Multiparadigm Perspectives on Theory Building. *Academy of Management Review*, 15(4), 584–602. <https://doi.org/10.5465/amr.1990.4310758>
- Gneiting, T., & Raftery, A. E. (2007). Strictly Proper Scoring Rules, Prediction, and Estimation. *Journal of the American Statistical Association*, 102(477), 359–378. <https://doi.org/10.1198/016214506000001437>
- Gold, A. H., Malhotra, A., & Segars, A. H. (2001). Knowledge Management: An Organizational Capabilities Perspective. *Journal of Management Information Systems*, 18(1), 185–214. <https://doi.org/10.1080/07421222.2001.11045669>
- Goleman, D. (1985). *Vital Lies, Simple Truths: The Psychology of Self Deception*. Simon and Schuster.
- Goodwin, P., Fildes, R., Lawrence, M., & Stephens, G. (2011). Restrictiveness and guidance in support systems. *Omega*, 39(3), 242–253. <https://doi.org/10.1016/j.omega.2010.07.001>

- Goodwin, P., Moritz, B., & Siemsen, E. (2018). Forecast Decisions. In *The Handbook of Behavioral Operations* (pp. 433–458). John Wiley & Sons, Ltd.
<https://doi.org/10.1002/9781119138341.ch12>
- Goodwin, P., & Wright, G. (2010). The limits of forecasting methods in anticipating rare events. *Technological Forecasting and Social Change*, 77(3), 355–368.
<https://doi.org/10.1016/j.techfore.2009.10.008>
- Gosling, S. D., Vazire, S., Srivastava, S., & John, O. P. (2004). Should We Trust Web-Based Studies? A Comparative Analysis of Six Preconceptions About Internet Questionnaires. *American Psychologist*, 59, 93–104. <https://doi.org/10.1037/0003-066X.59.2.93>
- Grant, R. M. (1996). Toward a knowledge-based theory of the firm. *Strategic Management Journal*, 17(S2), 109–122. <https://doi.org/10.1002/smj.4250171110>
- Hatch, M. J., & Schultz, M. (1997). Relations between Organizational Culture, Identity and Image. *European Journal of Marketing*, 31(5/6), 356–365. <https://doi.org/10.1108/eb060636>
- Henshel, R. L. (1993). Do self-fulfilling prophecies improve or degrade predictive accuracy? How sociology and economics can disagree and both be right. *The Journal of Socio-Economics*, 22(2), 85–104. [https://doi.org/10.1016/1053-5357\(93\)90017-F](https://doi.org/10.1016/1053-5357(93)90017-F)
- Herrigel, E., & Hull, R. F. C. (1953). *Zen in the art of archery*. Pantheon Books.
- Hesser, H. (2015). Modeling individual differences in randomized experiments using growth models: Recommendations for design, statistical analysis and reporting of results of internet interventions. *Internet Interventions*, 2(2), 110–120.
<https://doi.org/10.1016/j.invent.2015.02.003>
- Hoffman, A. J., & Ocasio, W. (2001). Not All Events Are Attended Equally: Toward a Middle-Range Theory of Industry Attention to External Events. *Organization Science*, 12(4), 414–434.
<https://doi.org/10.1287/orsc.12.4.414.10639>

- Hogarth, R. M., & Makridakis, S. (1981). Forecasting and Planning: An Evaluation. *Management Science*, 27(2), 115–138. <https://doi.org/10.1287/mnsc.27.2.115>
- Hollenbeck, J., Ilgen, D., Segó, D., Hedlund, J., Major, D., & Phillips, J. (1995). Multilevel Theory of Team Decision Making: Decision Performance in Teams Incorporating Distributed Expertise. *Journal of Applied Psychology*, 80(2), 292–316. <https://doi.org/10.1037/0021-9010.80.2.292>
- Ito, P. K. (1980). 7 Robustness of ANOVA and MANOVA test procedures. In *Handbook of Statistics* (Vol. 1, pp. 199–236). Elsevier. [https://doi.org/10.1016/S0169-7161\(80\)01009-7](https://doi.org/10.1016/S0169-7161(80)01009-7)
- Janis, I. L. (1982). *Groupthink: Psychological Studies of Policy Decisions and Fiascoes*. Houghton Mifflin.
- Janz, B. D., & Prasarnphanich, P. (2003). Understanding the Antecedents of Effective Knowledge Management: The Importance of a Knowledge-Centered Culture*. *Decision Sciences*, 34(2), 351–384. <https://doi.org/10.1111/1540-5915.02328>
- John, O. P., Donahue, E. M., & Kentle, R. L. (1991). *Big Five Inventory*. <https://doi.org/10.1037/t07550-000>
- Juul, S., Gluud, C., Simonsen, S., Frandsen, F. W., Kirsch, I., & Jakobsen, J. C. (2021). Blinding in randomised clinical trials of psychological interventions: A retrospective study of published trial reports. *BMJ Evidence-Based Medicine*, 26(3), 109–109. <https://doi.org/10.1136/bmjebm-2020-111407>
- Kahneman, D., & Lovallo, D. (1993). Timid Choices and Bold Forecasts: A Cognitive Perspective on Risk Taking. *Management Science*, 39(1), 17–31.
- Kahneman, D., & Tversky, A. (1979). Prospect Theory: An Analysis of Decision under Risk. *Econometrica*, 47(2), 263–291. JSTOR. <https://doi.org/10.2307/1914185>

- Kameda, T., Ohtsubo, Y., & Takezawa, M. (1997). Centrality in sociocognitive networks and social influence: An illustration in a group decision-making context. *Journal of Personality and Social Psychology, 73*(2), 296–309. <https://doi.org/10.1037/0022-3514.73.2.296>
- Kc, D., Staats, B. R., & Gino, F. (2013). Learning from My Success and from Others' Failure: Evidence from Minimally Invasive Cardiac Surgery. *Management Science, 59*(11), 2435–2449.
- Keith, N., & Frese, M. (2005). Self-Regulation in Error Management Training: Emotion Control and Metacognition as Mediators of Performance Effects. *Journal of Applied Psychology, 90*(4), 677–691. <https://doi.org/10.1037/0021-9010.90.4.677>
- Kerr, N. L., & Tindale, R. S. (2004). Group Performance and Decision Making. *Annual Review of Psychology, 55*(1), 623–655. <https://doi.org/10.1146/annurev.psych.55.090902.142009>
- Klayman, J. (1988). On the how and why (not) of learning from outcomes. In *Human judgment: The SJT view* (pp. 115–162). North-Holland. [https://doi.org/10.1016/S0166-4115\(08\)62172-X](https://doi.org/10.1016/S0166-4115(08)62172-X)
- Klein, G. (2008). Naturalistic Decision Making. *Human Factors, 50*(3), 456–460. <https://doi.org/10.1518/001872008X288385>
- Klein, G., Wiggins, S., & Dominguez, C. O. (2010). Team sensemaking. *Theoretical Issues in Ergonomics Science, 11*(4), 304–320. <https://doi.org/10.1080/14639221003729177>
- Kozlowski, S. W. J., & Chao, G. T. (2012). The Dynamics of Emergence: Cognition and Cohesion in Work Teams. *Managerial and Decision Economics, 33*(5–6), 335–354. <https://doi.org/10.1002/mde.2552>
- Kozlowski, S. W. J., & Chao, G. T. (2018). Unpacking team process dynamics and emergent phenomena: Challenges, conceptual advances, and innovative methods. *American Psychologist, 73*(4), 576–592. <https://doi.org/10.1037/amp0000245>

- Kozlowski, S. W. J., & Ilgen, D. R. (2006). Enhancing the Effectiveness of Work Groups and Teams. *Psychological Science in the Public Interest*, 7(3), 77–124.
<https://doi.org/10.1111/j.1529-1006.2006.00030.x>
- Kozlowski, S. W. J., & Ilgen, D. R. (2016). Enhancing the Effectiveness of Work Groups and Teams: *Psychological Science in the Public Interest*.
<https://journals.sagepub.com/doi/10.1111/j.1529-1006.2006.00030.x>
- Kremer, M., Moritz, B., & Siemsen, E. (2011). Demand Forecasting Behavior: System Neglect and Change Detection. *Management Science*, 57(10), 1827–1843.
<https://doi.org/10.1287/mnsc.1110.1382>
- Langer. (1998). *The Power of Mindful Learning*. Da Capo Books.
- Langer, E. J. (2000). Mindful learning. *Current Directions in Psychological Science*, 9(6), 220–223.
<https://doi.org/10.1111/1467-8721.00099>
- Langer, E. J. (2016). *The Power of Mindful Learning*. Hachette UK.
- Larey, T. S., & Paulus, P. B. (1999). Group Preference and Convergent Tendencies in Small Groups: A Content Analysis of Group Brainstorming Performance. *Creativity Research Journal*, 12(3), 175–184. https://doi.org/10.1207/s15326934crj1203_2
- Lawrence, M., Goodwin, P., O'Connor, M., & Önkál, D. (2006). Judgmental forecasting: A review of progress over the last 25years. *International Journal of Forecasting*, 22(3), 493–518.
<https://doi.org/10.1016/j.ijforecast.2006.03.007>
- Lawrence, M., & O'connor, M. (1995). The anchor and adjustment heuristic in time-series forecasting. *Journal of Forecasting*, 14(5), 443–451.
<https://doi.org/10.1002/for.3980140504>

- Lawrence, M., O'Connor, M., & Edmundson, B. (2000). A field study of sales forecasting accuracy and processes. *European Journal of Operational Research*, *122*(1), 151–160. [https://doi.org/10.1016/S0377-2217\(99\)00085-5](https://doi.org/10.1016/S0377-2217(99)00085-5)
- Lee, S. W. S. ., & Schwarz, N. (2014). Metaphor in judgment and decision making. In *The power of metaphor: Examining its influence on social life* (pp. 85–108). American Psychological Association. <https://doi.org/10.1037/14278-005>
- Lee, Y. S., & Siemsen, E. (2016). Task Decomposition and Newsvendor Decision Making. *Management Science*, *63*(10), 3226–3245. <https://doi.org/10.1287/mnsc.2016.2521>
- Legerstee, R., & Franses, P. H. (2014). Do Experts' SKU Forecasts Improve after Feedback? *Journal of Forecasting*, *33*(1), 69–79. <https://doi.org/10.1002/for.2274>
- Levinthal, D., & Rerup, C. (2006). Crossing an Apparent Chasm: Bridging Mindful and Less-Mindful Perspectives on Organizational Learning. *Organization Science*, *17*(4), 502–513. <https://doi.org/10.1287/orsc.1060.0197>
- Lewis, C. D. (Colin D. (n.d.). Industrial and business forecasting methods: A practical guide to exponential smoothing and curve fitting. (*No Title*). Retrieved July 9, 2023, from <https://cir.nii.ac.jp/crid/1130282269656766080>
- Linstone, H. A., & Turoff, M. (1975). *The Delphi method: Techniques and applications*. Addison-Wesley Pub. Co., Advanced Book Program.
- Lippman, S. A., & Rumelt, R. P. (1982). Uncertain Imitability: An Analysis of Interfirm Differences in Efficiency under Competition. *The Bell Journal of Economics*, *13*(2), 418–438. JSTOR. <https://doi.org/10.2307/3003464>
- Lovullo, D., Clarke, C., & Camerer, C. (2012). Robust analogizing and the outside view: Two empirical tests of case-based decision making. *Strategic Management Journal*, *33*(5), 496–512. <https://doi.org/10.1002/smj.962>

- Lovallo, D., & Kahneman, D. (2003). Delusions of success. How optimism undermines executives' decisions. *Harvard Business Review*, *81*(7), 56–63, 117.
- Madsen, P. M., & Desai, V. (2010). Failing to Learn? The Effects of Failure and Success on Organizational Learning in the Global Orbital Launch Vehicle Industry. *Academy of Management Journal*, *53*(3), 451–476. <https://doi.org/10.5465/amj.2010.51467631>
- Mahmoud, E., DeRoeck, R., Brown, R., & Rice, G. (1992). Bridging the gap between theory and practice in forecasting. *International Journal of Forecasting*, *8*(2), 251–267. [https://doi.org/10.1016/0169-2070\(92\)90123-Q](https://doi.org/10.1016/0169-2070(92)90123-Q)
- Maitlis, S., & Christianson, M. (2014). Sensemaking in Organizations: Taking Stock and Moving Forward. *The Academy of Management Annals*, *8*(1), 57–125. <https://doi.org/10.1080/19416520.2014.873177>
- Makridakis, S. G., Gaba, A., & Hogarth, R. (2010). *Dance with Chance: Making Luck Work for You*. Oneworld Publications.
- Makridakis, S., & Taleb, N. (2009). Decision making and planning under low levels of predictability. *International Journal of Forecasting*, *25*(4), 716–733. <https://doi.org/10.1016/j.ijforecast.2009.05.013>
- Makridakis, S., & Wheelwright, S. C. (1977). Forecasting: Issues & Challenges for Marketing Management. *Journal of Marketing*, *41*(4), 24–38. <https://doi.org/10.1177/002224297704100403>
- Marks, M. A., Zaccaro, S. J., & Mathieu, J. E. (2000). Performance implications of leader briefings and team-interaction training for team adaptation to novel environments. *Journal of Applied Psychology*, *85*(6), 971–986. <https://doi.org/10.1037/0021-9010.85.6.971>

- Mathieu, J. E., Heffner, T. S., Goodwin, G. F., Salas, E., & Cannon-Bowers, J. A. (2000). The influence of shared mental models on team process and performance. *Journal of Applied Psychology, 85*(2), 273–283. <https://doi.org/10.1037/0021-9010.85.2.273>
- McGahan, A. M. (2007). Academic Research That Matters to Managers: On Zebras, Dogs, Lemmings, Hammers, and Turnips. *Academy of Management Journal, 50*(4), 748–753. <https://doi.org/10.5465/amj.2007.26279166>
- McGrath, J. E., Arrow, H., & Berdahl, J. L. (2000). The Study of Groups: Past, Present, and Future. *Personality and Social Psychology Review, 4*(1), 95–105. https://doi.org/10.1207/S15327957PSPR0401_8
- Meehl, P. E. (1954). *Clinical versus statistical prediction: A theoretical analysis and a review of the evidence* (pp. x, 149). University of Minnesota Press. <https://doi.org/10.1037/11281-000>
- Mellers, B., Ungar, L., Baron, J., Ramos, J., Gurcay, B., Fincher, K., Scott, S. E., Moore, D., Atanasov, P., Swift, S. A., Murray, T., Stone, E., & Tetlock, P. E. (2014). Psychological Strategies for Winning a Geopolitical Forecasting Tournament. *Psychological Science, 25*(5), 1106–1115. <https://doi.org/10.1177/0956797614524255>
- Mello, J. (2009). The Impact of Sales Forecast Game Playing on Supply Chains. *Foresight: The International Journal of Applied Forecasting, 13*, 13–22.
- Mentzer, J. T., Flint, D. J., & Kent, J. L. (1999). Developing a logistics service quality scale. *Journal of Business Logistics, 20*(1), 9.
- Meyer, M. W., & Zucker, L. G. (1989). *Permanently Failing Organizations*. SAGE Publications.
- Mezias, J. M., & Starbuck, W. H. (2003). Studying the Accuracy of Managers' Perceptions: A Research Odyssey. *British Journal of Management, 14*(1), 3–17. <https://doi.org/10.1111/1467-8551.00259>

- Mezias, J., & Starbuck, W. H. (2009). Decision Making with Inaccurate, Unreliable Data. *The Oxford Handbook of Organizational Decision Making*.
<https://doi.org/10.1093/oxfordhb/9780199290468.003.0004>
- Miller, D. (1994). What Happens After Success: The Perils of Excellence*. *Journal of Management Studies*, 31(3), 325–358. <https://doi.org/10.1111/j.1467-6486.1994.tb00621.x>
- Moon, M. A., Mentzer, J. T., & Smith, C. D. (2003). Conducting a sales forecasting audit. *International Journal of Forecasting*, 19(1), 5–25. [https://doi.org/10.1016/S0169-2070\(02\)00032-8](https://doi.org/10.1016/S0169-2070(02)00032-8)
- Morgeson, F. P., & Hofmann, D. A. (1999). The Structure and Function of Collective Constructs: Implications for Multilevel Research and Theory Development. *The Academy of Management Review*, 24(2), 249–265. JSTOR. <https://doi.org/10.2307/259081>
- Morrison, J. E., & Meliza, L. L. (1999). *Foundations of the After Action Review Process* (IDA/HQ-D2332). INSTITUTE FOR DEFENSE ANALYSES ALEXANDRIA VA.
<https://apps.dtic.mil/docs/citations/ADA368651>
- Mosier, K., Fischer, U., Hoffman, R. R., & Klein, G. (2018). Expert professional judgments and “naturalistic decision making.” In *The Cambridge handbook of expertise and expert performance*, 2nd ed (pp. 453–475). Cambridge University Press.
<https://doi.org/10.1017/9781316480748.025>
- Mullen, B., Johnson, C., & Salas, E. (1991). Productivity loss in brainstorming groups: A meta-analytic integration. *Basic and Applied Social Psychology*, 12(1), 3–23.
https://doi.org/10.1207/s15324834basp1201_1
- Muller, K. E., Lavange, L. M., Ramey, S. L., & Ramey, C. T. (1992). Power Calculations for General Linear Multivariate Models Including Repeated Measures Applications. *Journal of the*

- American Statistical Association*, 87(420), 1209–1226.
<https://doi.org/10.1080/01621459.1992.10476281>
- Nijstad, B. A., & Stroebe, W. (2006). How the Group Affects the Mind: A Cognitive Model of Idea Generation in Groups. *Personality and Social Psychology Review*, 10(3), 186–213.
https://doi.org/10.1207/s15327957pspr1003_1
- Nonaka, I. (1994). A Dynamic Theory of Organizational Knowledge Creation. *Organization Science*, 5(1), 14–37. <https://doi.org/10.1287/orsc.5.1.14>
- Nonaka, I., Byosiere, P., Borucki, C. C., & Konno, N. (1994). Organizational knowledge creation theory: A first comprehensive test. *International Business Review*, 3(4), 337–351.
[https://doi.org/10.1016/0969-5931\(94\)90027-2](https://doi.org/10.1016/0969-5931(94)90027-2)
- Ocasio, W. (1997). Towards an Attention-Based View of the Firm. *Strategic Management Journal*, 18(S1), 187–206. [https://doi.org/10.1002/\(SICI\)1097-0266\(199707\)18:1+<187::AID-SMJ936>3.0.CO;2-K](https://doi.org/10.1002/(SICI)1097-0266(199707)18:1+<187::AID-SMJ936>3.0.CO;2-K)
- O’Dell, C., & Grayson, C. J. (1998). If Only We Knew What We Know: Identification and Transfer of Internal Best Practices. *California Management Review*, 40(3), 154–174.
<https://doi.org/10.2307/41165948>
- O’Dell, C. S. (2011). *The New Edge in Knowledge: How Knowledge Management is Changing the Way We Do Business*. Wiley.
- Oliva, R., & Watson, N. (2009). Managing Functional Biases in Organizational Forecasts: A Case Study of Consensus Forecasting in Supply Chain Planning. *Production and Operations Management*, 18(2), 138–151. <https://doi.org/10.1111/j.1937-5956.2009.01003.x>
- Oliva, R., & Watson, N. (2011). Cross-functional alignment in supply chain planning: A case study of sales and operations planning. *Journal of Operations Management*, 29(5), 434–448.
<https://doi.org/10.1016/j.jom.2010.11.012>

- Parikh, M., Fazlollahi, B., & Verma, S. (2001). The Effectiveness of Decisional Guidance: An Empirical Evaluation. *Decision Sciences*, 32(2), 303–332. <https://doi.org/10.1111/j.1540-5915.2001.tb00962.x>
- Patriotta, G. (2003). Sensemaking on the Shop Floor: Narratives of Knowledge in Organizations*. *Journal of Management Studies*, 40(2), 349–375. <https://doi.org/10.1111/1467-6486.00343>
- Peer, E., Brandimarte, L., Samat, S., & Acquisti, A. (2017). Beyond the Turk: Alternative platforms for crowdsourcing behavioral research. *Journal of Experimental Social Psychology*, 70, 153–163. <https://doi.org/10.1016/j.jesp.2017.01.006>
- Penrose, E., Pitelis, C., Penrose, E., & Pitelis, C. (2009). *The Theory of the Growth of the Firm* (Fourth Edition, Fourth Edition). Oxford University Press.
- Pettigrew, A. M. (1979). On Studying Organizational Cultures. *Administrative Science Quarterly*, 24(4), 570–581. JSTOR. <https://doi.org/10.2307/2392363>
- Polanyi, M. (2009). *The Tacit Dimension*. University of Chicago Press.
- Poole, M. S., Seibold, D. R., & McPhee, R. D. (1985). Group decision-making as a structural process. *Quarterly Journal of Speech*, 71(1), 74–102. <https://doi.org/10.1080/00335638509383719>
- Poole, M. S., Seibold, D. R., & McPhee, R. D. (1996). The structuration of group decisions. In *Communication and group decision making, 2nd ed* (pp. 114–146). Sage Publications, Inc. <https://doi.org/10.4135/9781452243764.n5>
- Protzner, S., & Velde, S. van de. (2015). *Mind the gap between demand and supply: A behavioral perspective on demand forecasting*.

- Rammstedt, B., & John, O. P. (2007). Measuring personality in one minute or less: A 10-item short version of the Big Five Inventory in English and German. *Journal of Research in Personality, 41*(1), 203–212. <https://doi.org/10.1016/j.jrp.2006.02.001>
- Ray, J. L., Baker, L. T., & Plowman, D. A. (2011). Organizational Mindfulness in Business Schools. *Academy of Management Learning & Education, 10*(2), 188–203. <https://doi.org/10.5465/amle.10.2.zqr188>
- Renecle, M., Tomás, I., Gracia, F. J., & Peiró, J. M. (2020). Spanish validation of the mindful organizing scale: A questionnaire for the assessment of collective mindfulness. *Accident Analysis & Prevention, 134*, 105351. <https://doi.org/10.1016/j.aap.2019.105351>
- Ron, N., Lipshitz, R., & Popper, M. (2006). How Organizations Learn: Post-flight Reviews in an F-16 Fighter Squadron. *Organization Studies, 27*(8), 1069–1089. <https://doi.org/10.1177/0170840606064567>
- Rosemberg, E. (2023). Forecasting in Organizations: Reinterpreting Collective Judgment Through Mindful Organizing. *International Series in Operations Research & Management Science, 289–309*.
- Rowe, G., & Wright, G. (1996). The impact of task characteristics on the performance of structured group forecasting techniques. *International Journal of Forecasting, 12*(1), 73–89. [https://doi.org/10.1016/0169-2070\(95\)00658-3](https://doi.org/10.1016/0169-2070(95)00658-3)
- Rowe, G., & Wright, G. (2001). Expert Opinions in Forecasting: The Role of the Delphi Technique. In J. S. Armstrong (Ed.), *Principles of Forecasting: A Handbook for Researchers and Practitioners* (pp. 125–144). Springer US. https://doi.org/10.1007/978-0-306-47630-3_7
- Sagan, S. D. (1995). *The Limits of Safety: Organizations, Accidents, and Nuclear Weapons*. Princeton University Press.

- Salas, E., Cannon-Bowers, J. A., & Johnston, J. H. (1998). Lessons learned from conducting the TADMUS program: Balancing science, practice, and more. In *Making decisions under stress: Implications for individual and team training* (pp. 409–413). American Psychological Association. <https://doi.org/10.1037/10278-017>
- Scheele, L. M., Thonemann, U. W., & Slikker, M. (2018). Designing Incentive Systems for Truthful Forecast Information Sharing Within a Firm. *Manage. Sci.*, *64*(8), 3690–3713. <https://doi.org/10.1287/mnsc.2017.2805>
- Schein, E. (1990). Organizational Culture. *American Psychologist*, *45*(2), 109–119.
- Schleppenbach, M., Flevares, L. M., Sims, L. M., & Perry, M. (2007). Teachers' Responses to Student Mistakes in Chinese and U.S. Mathematics Classrooms. *The Elementary School Journal*, *108*(2), 131–147. JSTOR. <https://doi.org/10.1086/525551>
- Schneider, B., Ehrhart, M. G., & Macey, W. H. (2013). Organizational Climate and Culture. *Annual Review of Psychology*, *64*(1), 361–388. <https://doi.org/10.1146/annurev-psych-113011-143809>
- Schneider, B., & Reichers, A. E. (1983). On the Etiology of Climates. *Personnel Psychology*, *36*(1), 19–39. <https://doi.org/10.1111/j.1744-6570.1983.tb00500.x>
- Sears, D. O. (1986). College sophomores in the laboratory: Influences of a narrow data base on social psychology's view of human nature. *Journal of Personality and Social Psychology*, *51*, 515–530. <https://doi.org/10.1037/0022-3514.51.3.515>
- Shapira, Z. (2008, March 6). *On The Implications of Behavioral Decision Theory for Managerial Decision Making: Contributions and Challenges*. The Oxford Handbook of Organizational Decision Making. <https://doi.org/10.1093/oxfordhb/9780199290468.003.0015>
- Sitkin, S. B. (1992). Learning Through Failure: The Strategy of Small Losses. *Research in Organizational Behavior*, *14*, 231–266.

- SkillsPacks (Director). (2015, July 23). *Leadership Styles: Which Type of Leader Are You?*
https://www.youtube.com/watch?v=ddt_IGMMOrI
- Slovic, P., Fischhoff, B., & Lichtenstein, S. (1977). Behavioral Decision Theory. *Annual Review of Psychology*, 28(1), 1–39. <https://doi.org/10.1146/annurev.ps.28.020177.000245>
- Slovic, P., & Lichtenstein, S. (1971). Comparison of Bayesian and regression approaches to the study of information processing in judgment. *Organizational Behavior and Human Performance*, 6(6), 649–744. [https://doi.org/10.1016/0030-5073\(71\)90033-X](https://doi.org/10.1016/0030-5073(71)90033-X)
- Smith-Jentsch, K. A., Cannon-Bowers, J. A., Tannenbaum, S. I., & Salas, E. (2008). Guided team self-correction: Impacts on team mental models, processes, and effectiveness. *Small Group Research*, 39(3), 303–327. <https://doi.org/10.1177/1046496408317794>
- Sniezek, J. A. (1989). An examination of group process in judgmental forecasting. *International Journal of Forecasting*, 5(2), 171–178. [https://doi.org/10.1016/0169-2070\(89\)90084-8](https://doi.org/10.1016/0169-2070(89)90084-8)
- Starbuck, W. H. (2015). Karl E. Weick and the dawning awareness of organized cognition. *Management Decision*, 53(6), 1287–1299. <https://doi.org/10.1108/MD-04-2014-0183>
- Staw, B. M., Sandelands, L. E., & Dutton, J. E. (1981). Threat Rigidity Effects in Organizational Behavior: A Multilevel Analysis. *Administrative Science Quarterly*, 26(4), 501–524. JSTOR. <https://doi.org/10.2307/2392337>
- Stevenson, H. W., & Stigler, J. W. (1992). *The learning gap: Why our schools are failing and what we can learn from Japanese and Chinese education*. Summit Books.
- Stewart, T. R. (2001). Improving Reliability of Judgmental Forecasts. In J. S. Armstrong (Ed.), *Principles of Forecasting: A Handbook for Researchers and Practitioners* (pp. 81–106). Springer US. https://doi.org/10.1007/978-0-306-47630-3_5
- Sunstein, C. R. (2015). *Wiser: Getting beyond groupthink to make better decisions*. Harvard Business Review Press.

- Surowiecki, J. (2005). *The Wisdom of Crowds*. Anchor.
- Sutcliffe, K. M. (1994). What Executives Notice: Accurate Perceptions in Top Management Teams. *The Academy of Management Journal*, 37(5), 1360–1378. JSTOR. <https://doi.org/10.2307/256677>
- Sutcliffe, K. M., Vogus, T. J., & Dane, E. (2016). Mindfulness in Organizations: A Cross-Level Review. *Annual Review of Organizational Psychology and Organizational Behavior*, 3(1), 55–81. <https://doi.org/10.1146/annurev-orgpsych-041015-062531>
- Sutcliffe, K. M., & Weber, K. (2003). The High Cost of Accurate Knowledge. *Harvard Business Review*, 81(5), 74–82.
- Szulanski, G. (1996). Exploring Internal Stickiness: Impediments to the Transfer of Best Practice Within the Firm. *Strategic Management Journal*, 17, 27–43. JSTOR.
- Tannenbaum, S. I., & Cerasoli, C. P. (2013). Do Team and Individual Debriefs Enhance Performance? A Meta-Analysis. *Human Factors*, 55(1), 231–245. <https://doi.org/10.1177/0018720812448394>
- Tannenbaum, S. I., Smith-Jentsch, K. A., & Behson, S. J. (1998). Training team leaders to facilitate team learning and performance. In *Making decisions under stress: Implications for individual and team training* (pp. 247–270). American Psychological Association. <https://doi.org/10.1037/10278-009>
- Taylor, S. E. (1989). *Positive illusions: Creative self-deception and the healthy mind*. Basic Books.
- Teece, D. J. (2004). Knowledge and Competence as Strategic Assets. In C. W. Holsapple (Ed.), *Handbook on Knowledge Management 1: Knowledge Matters* (pp. 129–152). Springer. https://doi.org/10.1007/978-3-540-24746-3_7
- Tetlock, P. E., & Gardner, D. (2015). *Superforecasting: The art and science of prediction* (p. 340). Crown Publishers/Random House.

- Tindale, R. S. (1993). Decision errors made by individuals and groups. In *Individual and group decision making: Current issues* (pp. 109–124). Lawrence Erlbaum Associates, Inc.
- Trapmann, S., Hell, B., Hirn, J.-O. W., & Schuler, H. (2007). Meta-Analysis of the Relationship Between the Big Five and Academic Success at University. *Zeitschrift Für Psychologie / Journal of Psychology*, *215*(2), 132–151. <https://doi.org/10.1027/0044-3409.215.2.132>
- Tversky, A., & Kahneman, D. (1974). Judgment under Uncertainty: Heuristics and Biases. *Science*, *185*(4157), 1124–1131. <https://doi.org/10.1126/science.185.4157.1124>
- van Dyck, C., Frese, M., Baer, M., & Sonnentag, S. (2005). Organizational Error Management Culture and Its Impact on Performance: A Two-Study Replication. *Journal of Applied Psychology*, *90*(6), 1228–1240. <https://doi.org/10.1037/0021-9010.90.6.1228>
- Vaughan, D. (1996). *The Challenger Launch Decision: Risky Technology, Culture, and Deviance at NASA*. University of Chicago Press.
- Vogus, T., & Colville, I. (2017). Sensemaking, simplicity and mindfulness. *The Sage Handbook of Process Organization Studies*, 340–355.
- Vogus, T. J. (2011). *Mindful Organizing: Establishing and Extending the Foundations of Highly Reliable Performance* (SSRN Scholarly Paper ID 1904613). Social Science Research Network. <https://papers.ssrn.com/abstract=1904613>
- Vogus, T. J., Cooil, B., Sitterding, M., & Everett, L. Q. (2014). Safety Organizing, Emotional Exhaustion, and Turnover in Hospital Nursing Units. *Medical Care*, *52*(10), 870–876. JSTOR.
- Vogus, T. J., & Sutcliffe, K. M. (2007). The Safety Organizing Scale: Development and Validation of a Behavioral Measure of Safety Culture in Hospital Nursing Units. *Medical Care*, *45*(1), 46–54. JSTOR.

- Vogus, T. J., & Sutcliffe, K. M. (2012). Organizational Mindfulness and Mindful Organizing: A Reconciliation and Path Forward. *Academy of Management Learning & Education*, 11(4), 722–735. <https://doi.org/10.5465/amle.2011.0002c>
- Wang, G., Oh, I.-S., Courtright, S. H., & Colbert, A. E. (2011). Transformational Leadership and Performance Across Criteria and Levels: A Meta-Analytic Review of 25 Years of Research. *Group & Organization Management*, 36(2), 223–270. <https://doi.org/10.1177/1059601111401017>
- Wang, T., & Murphy, J. (2004). An Examination of Coherence in a Chinese Mathematics Classroom. In *How Chinese Learn Mathematics: Vol. Volume 1* (pp. 107–123). WORLD SCIENTIFIC. https://doi.org/10.1142/9789812562241_0004
- Weick, K. E. (1988). Enacted Sensemaking in Crisis Situations[1]. *Journal of Management Studies*, 25(4), 305–317. <https://doi.org/10.1111/j.1467-6486.1988.tb00039.x>
- Weick, K. E. (1995). *Sensemaking in organizations*. Sage Publications.
- Weick, K. E., & Roberts, K. H. (1993). Collective Mind in Organizations: Heedful Interrelating on Flight Decks. *Administrative Science Quarterly*, 38(3), 357–381. JSTOR. <https://doi.org/10.2307/2393372>
- Weick, K. E., Sutcliffe, K. M., & Obstfeld, D. (1999). Organizing for high reliability: Processes of collective mindfulness. In *Research in organizational behavior*, Vol. 21 (pp. 81–123). Elsevier Science/JAI Press.
- Weick, K. E., Sutcliffe, K. M., & Obstfeld, D. (2005). Organizing and the Process of Sensemaking. *Organization Science*, 16(4), 409–421. <https://doi.org/10.1287/orsc.1050.0133>
- Weick, K., & Sutcliffe, K. (2007). *Managing the unexpected: Resilient performance in an age of uncertainty, second edition*. John Wiley & Sons, Inc.

Winklhofer, H., Diamantopoulos, A., & Witt, S. F. (1996). Forecasting practice: A review of the empirical literature and an agenda for future research. *International Journal of Forecasting*, *12*(2), 193–221. [https://doi.org/10.1016/0169-2070\(95\)00647-8](https://doi.org/10.1016/0169-2070(95)00647-8)

Zhou, H., & Fishbach, A. (2016). The pitfall of experimenting on the web: How unattended selective attrition leads to surprising (yet false) research conclusions. *Journal of Personality and Social Psychology*, *111*(4), 493–504. <https://doi.org/10.1037/pspa0000056>

Zollo, M., & Winter, S. G. (2002). Deliberate Learning and the Evolution of Dynamic Capabilities. *Organization Science*, *13*(3), 339–351. <https://doi.org/10.1287/orsc.13.3.339.2780>