IE UNIVERSIDAD

TESIS DOCTORAL / DOCTORAL DISSERTATION

TRES ENSAYOS SOBRE INNOVACIÓN Y CREATIVIDAD ORGANIZACIONAL UTILIZANDO ENFOQUES BASADOS EN TEXTO/ THREE ESSAYS ON ORGANIZATIONAL INNOVATION AND CREATIVITY USING TEXT-BASED APPROACHES.

CHANDRIKA RATHEE

March, 2023

IE UNIVERSIDAD

TESIS DOCTORAL / DOCTORAL DISSERTATION

TRES ENSAYOS SOBRE INNOVACIÓN Y CREATIVIDAD ORGANIZACIONAL UTILIZANDO ENFOQUES BASADOS EN TEXTO/ THREE ESSAYS ON ORGANIZATIONAL INNOVATION AND CREATIVITY USING TEXT-BASED APPROACHES.

CHANDRIKA RATHEE

DOCTORAL THESIS ADVISORS:

MARCO GIARRATANA, Ph.D.

ABSTRACT

This dissertation aims at presenting some contextual factors that influence innovation, including economic, institutional, and cultural factors, which are critical for understanding the complex and dynamic environment in which innovation occurs. It consists of three empirical studies that leverage novel text-based measures and methodological advancements to generate new insights to inform innovation theory. Chapter one explores how to synthesize innovation knowledge generated from academic and industrial endeavors by analyzing the language used in patent documents and scientific publications. Chapter two investigates how institutional interventions modify the competitive space of firms and affect their disclosure behavior by studying the impact of perceived competition on the dissemination of innovation or R&D findings on a regulatory platform. Finally, chapter three examines how dynastic embeddedness enables and constrains innovation production by developing theory and testing the relationship between the dynastic roots of producers and creative productions.

RESUMEN

Esta tesis presenta un estudio sobre los factores contextuales que influyen en la innovación, incluidos los factores económicos, institucionales y culturales, que son fundamentales para comprender el entorno complejo y dinámico en el que se produce la innovación. Incluye tres estudios empíricos que aprovechan medidas novedosas basadas en textos y avances metodológicos para generar nuevos conocimientos que sirvan de base a la teoría de la innovación. El capítulo uno explora cómo sintetizar el conocimiento de la innovación generado a partir de esfuerzos académicos e industriales mediante el análisis del lenguaje utilizado en documentos de patentes y publicaciones científicas. El capítulo dos investiga cómo las intervenciones institucionales modifican el espacio competitivo de las empresas y afectan a su comportamiento de divulgación estudiando el impacto de la competencia percibida en la difusión de los hallazgos de innovación o I+D en una plataforma reguladora. Por último, el capítulo tres examina cómo el arraigo dinástico posibilita y limita la producción de innovación mediante el desarrollo de una teoría y la comprobación de la relación entre el arraigo dinástico de los productores y las producciones creativas.

ACKNOWLEDGEMENTS

This thesis is a culmination of the immense generosity, invaluable assistance, and boundless inspiration from an array of named and unnamed colleagues and mentors who have contributed significantly to shaping my academic journey. I am profoundly grateful to Dr. Marco Giarratana, whose unwavering confidence in my ideas and patient guidance has been instrumental in developing this work. His wealth of experience and deep insights have been a constant source of inspiration and motivation for me to strive for excellence as a researcher. I also express my heartfelt gratitude to Dr. Rossella Salandra for her unwavering support and for emphasizing the importance of ideas, camaraderie, and community in academia. She embodies the very best qualities of a scholar and has been an exceptional role model. Dr. Oliver Alexy's contributions have been invaluable in shaping my thinking process and instilling in me the value of asking the right questions. From him, I learned the subtleties necessary to enhance the impact of my research. Dr. Matthias Seifert sparked my interest in quantitative research methods and helped me realize that being an introvert can be a valuable asset in academia. Dr. Luis Diestre has been an inspirational figure and instrumental in honing my critical thinking skills. His strategy seminar helped me develop and practice techniques to purposefully critique and evaluate the value and limitations of ideas, a crucial skill for any researcher.

My gratitude extends beyond my dissertation committee members to my numerous coauthors, colleagues, my Ph.D. cohort at IE, and all the faculty members at IE Business School for their unrelenting support and inspiration, which helped me strive to achieve my full potential. I am indebted to all of my coauthors and colleagues for their unrelenting support, without which I would not have been able to strive toward my full potential. I am particularly grateful to Dr. Konstantina Valogianni for being a prolific, prompt, and practical coauthor and for inspiring me with her conduct. And to Dr. Luis Gomez-Mejia, thank you for teaching the value of humility and simplicity in approach, which I believe is the key to your tremendous success as a distinguished scholar. I have learned something valuable from every interaction with Dr. Julien Jourdan and Dr. Taiyuan Wang. Their perspectives and approaches to the scholarly conversation have been a practical learning experience. And to all of the IE faculty members, including Dr. Juan Santaló, Dr. Caterina Moschieri, Dr. Martina Montauti, Dr. Manos Avgerinos, Dr. Mikko Ketokivi, Dr. David Gómez-Ullate Oteiza, and Dr. Kriti Jain, thank you for sharing your expertise and making such a profound impact on my research outlook. These exceptional individuals have profoundly impacted my research outlook and shaped my conduct as a researcher in subtle ways.

I am also incredibly grateful to IE University, not just for the financial aid and opportunity to pursue my Ph.D., but for the rich cultural and research environment that has made this journey such a rewarding learning experience. I am also indebted to the Ph.D. office for facilitating our scholarly endeavors and looking after our welfare at IE. I especially want to thank Dr. Julio de Castro and Dr. Laura Maguire for their unwavering support throughout my scholarly journey.

Above all, I dedicate this work to my beloved grandmother, Sukhshanti Devi, whose stories full of cultural insights intrigued me and instilled in me a lifelong search for truth and wisdom. My parents, Kamlesh Rathee and Mahender Singh Rathee provided me with a nurturing and safe space to dream and aspire. My siblings, Monica, Arun, and Ranjay, enriched my perspective, forcing me to be my best self. They have been a constant source of love, camaraderie, and belonging. This thesis would not have been possible without Dr. Sumeet Malik, my husband, whose support, mentorship, and critiques pushed me to my limits by setting a high bar for me to match as a researcher. My children, Nikita and Theo, have been a source of love and divine grace. Nikita witnessed my Ph.D. journey and provided priceless perspectives on my ideas, making thinking fun and simple.

Lastly, I want to thank all the colleagues and comrades who have interacted with my research and ideas during numerous conference reviews and presentations at various forums, including the Academy of Management, DRUID, Strategic Management Society, European Group for Organizational Studies, and special seminars hosted by Bologna Business School, Rotterdam School of Management and Bath School of Management. They have knowingly or unknowingly instilled and helped polish some parts of me as a thinker, researcher, and presenter. And finally, I acknowledge the entire body of researchers, philosophers, and poets whose works inspired the development of my thesis and me as a researcher. As they say, we stand on the shoulders of giants. Indeed, without prior groundbreaking work and insights, this thesis would not have been possible.

I offer my sincere thanks to all of you, named or unnamed. You've all been essential to my academic journey and made this thesis possible.

TABLE OF CONTENTS

INTRODUCTION

Innovation—technological or creative—is the most studied concept in organizational studies, given its colossal implication on a firm's competitive advantage and performance in an increasingly changing environment (Dess & Picken, 2000; Tushman & O'Reilly, 1996). Innovation's importance in driving growth, competitive advantage, and the overall success of organizations cannot be overstated. A keyword search of "organizational innovation" on Google Scholar yields thousands of articles addressing various dimensions of organizational innovation and creativity, indicating its importance and complexity in the academic discourse.

Given its multidisciplinary impact, the subject of innovation has been studied across various branches of science, engineering, humanities, arts, and social sciences, making it a highly multidimensional concept with various connotations. It is studied at different levels of analysis, including person, group, organization, industry, and economy. Godin (2006) identifies over ten related conceptualizations of innovation historically used to portray innovation over time. This dissertation explores internal and external characteristics that help synthesize, share, and shape organizational innovation. Innovation is not a one-size-fits-all process and is shaped by the unique characteristics of the environment in which it occurs (Tödtling & Trippl, 2005). Therefore studying contextual factors in innovation research is critical for understanding the complex and dynamic environment in which innovation occurs (Kimberly & Evanisko, 1981). Contextual factors refer to the external conditions and internal circumstances that influence the innovation process, including economic, institutional, and cultural elements. To put it poetically—context is to creativity and innovation what soil is to the seed. Thus, studying innovation by embracing contextual richness might help gain a deeper understanding of innovation processes and to generate theoretical insights. This knowledge is crucial to inform the development of strategies and policies supporting and promoting innovation.

Recent advancements in computer science technologies, such as machine learning and natural language processing (NLP), have enabled new methods to understand and measure innovation. In a short time, a substantial body of research has emerged on exploring the meanings embedded in the organizational innovation artifacts such as patents, scientific publications, and product attributes (Becker, Rullani, & Zirpoli, 2021; Bellstam, Bhagat, & Cookson, 2021; Boudreau, Guinan, Lakhani, & Riedl, 2016; Goldberg, Hannan, & Kovács, 2016; Hannigan et al., 2019; Vakili & Kaplan, 2021; Verhoeven, Bakker, & Veugelers, 2016; Veugelers & Wang, 2019). Capturing subtle linguistic patterns and improved construct representations via computer-based language processing tools have amplified and sharpened the linguistic turn in innovation research (Alvesson & Kärreman, 2000), enabling the exploration of cognitive underpinnings of innovation (DiMaggio, 2015). Also, the exponential leap in information access and processing capabilities driving the big data revolution has made information a hot commodity (Khan & Vorley, 2017; McAfee, Brynjolfsson, Davenport, Patil, & Barton, 2012). For instance, it is now possible to aggregate and synthesize the inherent knowledge maps of research and product development (Suominen, Toivanen, & Seppänen, 2017). These advancements affect every aspect of the research and innovation process, from idea generation to the diffusion of innovation, and have an invigorating effect on organizational studies and innovation theory. This dissertation supports the view that Innovation theory can be strengthened by exploring the subtle organizational learning and innovation mechanisms made possible by the enhanced ability to analyze tonal patterns of fine-grained data (Joseph, Rhee, & Wilson, 2022) and tries to leverage these advancements to study three interesting phenomenon.

The three empirical studies are focused on understanding innovation production, sharing, and synthesis. These studies are situated in high technology and creative contexts and leverage text-based measures and methodological advancements to generate new insights to inform innovation theory. Chapter one devises a methodology to synthesis and combine knowledge from large bodies of patents and academic publications texts with an aim to assess the value of patented technology. The second chapter explores the impact of perceived competition on the dissemination of innovation or R&D findings on a regulatory platform, including the tonal variations of information. The third chapter is situated in the unique and rich context of Bollywood- Hindi film industry, which is dominated by dynastic producers. The study develops theory and tests the relationship between the dynastic roots of producers and creative productions.

The first chapter, “Uncovering The Performance Targets Of Patented Technology Using A Blended Approach Of Topic Modeling And K-Nearest Neighbor Algorithm”, devises a blended approach that maps the knowledge disclosed in scientific publications to patented inventions to uncover performance targets. We apply our technique to the Electric Energy Storage (EES) research for electric vehicles and show how it can reveal the performance targets not explicitly stated in the patent's text. Our blended algorithm performs significantly better than topic modeling on patents alone in assessing thematic similarity. Our study highlights the usefulness of blending alternate information sources with different disclosure levels to uncover performance targets of patented technology, particularly in complex component-based technologies.

The second chapter, titled "Regulatory disclosures and competitive signals: Competition affecting speed and tenor of clinical trial results," argues that regulatory disclosure mandates requiring firms to share information can exacerbate competitive pressures. However, the extant literature views regulatory compliance as a burden for firms that abide reluctantly, overlooking competitive dynamics and signaling. Combining insights from the literature on regulatory disclosures and competitive signaling theory, we propose that when regulatory disclosure mandates pit firms against each other, they respond to the magnified competitive pressures by swift information sharing and adopting a more positive tenor in their disclosures. Furthermore, firms with a reputation for introducing novel products display more urgency and fervor to signal to competitors and reduce uncertainty about their ongoing product developments. We test our hypotheses in the pharmaceutical setting, leveraging a multi-source dataset comprising clinical trials completed between 2008 and 2019 and registered on ClinicalTrials.gov under the ambit of FDAAA 801 regulation, information on sponsoring firms from Compustat and the FDA Orange Book, and the publications linked to the trials extracted from PubMed.

The third chapter, "The dynastic duality: Examining the sociocultural capital and nepotism discount of Bollywood dynasts," explores the product atypicality of family-run organizations in a creative cultural context. Family producers are the most prevalent organizational form globally. They are driven by the dynastic motive—to perpetuate their wealth and status over generations so that the descendants can enjoy the fruits of their forebears' labor. However, there are conflicting views on whether descendants of these family producers are innovative or constrained by their ancestors' legacies. On the one hand, descendants may inherit status and tacit knowledge from their ancestors. Thus, they enjoy the benefits of accessing and bundling coveted resources, allowing them to pursue innovation through tradition. On the other hand, descendants may be constrained by skepticism about nepotism and the unwillingness of the controlling families to professionalize. In this study, we attempt to understand the dynasty's role in family producers to innovate. We show that the dynastic roots of producers in Bollywood (Hindi movie industry) can lead to an inverted U-shaped innovation curve, where sociocultural capital and nepotism discount can both hinder or enable innovation.

The three chapters offer new and focused perspectives on essential elements of innovation, namely synthesis, disclosure, and production within the studied contexts. The studies aim to construct and test theories about combining knowledge from academic and industry sources, the impact of competition on sharing innovation, and the facilitating or limiting role of family connections on creativity and innovations.

In conclusion, this dissertation adds to the literature on organizational innovation, specifically to the research addressing innovation production, sharing, and synthesis. The studies presented provide valuable insights for scholars and practitioners discussed in detail in each chapter and suggest ways for future research in the studied innovation streams.

REFERENCES

Alvesson, M., & Kärreman, D. 2000. Taking the Linguistic Turn in Organizational Research: Challenges, Responses, Consequences. The Journal of Applied Behavioral Science, 36(2): 136–158.

Becker, M. C., Rullani, F., & Zirpoli, F. 2021. The role of digital artefacts in early stages of distributed innovation processes. Research Policy, 50(10): 104349.

Bellstam, G., Bhagat, S., & Cookson, J. A. 2021. A text-based analysis of corporate innovation. Management Science, 67(7): 4004–4031.

Boudreau, K. J., Guinan, E. C., Lakhani, K. R., & Riedl, C. 2016. Looking Across and Looking Beyond the Knowledge Frontier: Intellectual Distance, Novelty, and Resource Allocation in Science. Management Science, 62(10): 2765–2783.

Dess, G. G., & Picken, J. C. 2000. Changing roles: Leadership in the 21st century. Organizational Dynamics, 28(3): 18–34.

DiMaggio, P. 2015. Adapting computational text analysis to social science (and vice versa). Big Data & Society, 2(2): 2053951715602908.

Godin, B. 2006. The Linear Model of Innovation: The Historical Construction of an Analytical Framework. Science, Technology, & Human Values, 31(6): 639–667.

Goldberg, A., Hannan, M. T., & Kovács, B. 2016. What Does It Mean to Span Cultural Boundaries? Variety and Atypicality in Cultural Consumption. American Sociological Review, 81(2): 215–241.

Hannigan, T. R., Haans, R. F., Vakili, K., Tchalian, H., Glaser, V. L., et al. 2019. Topic modeling in management research: Rendering new theory from textual data. Academy of Management Annals, 13(2): 586–632.

Joseph, J., Rhee, L., & Wilson, A. J. 2022. Corporate Hierarchy and Organizational Learning: Member Turnover, Code Change, and Innovation in the Multiunit Firm. Organization Science. https://doi.org/10.1287/orsc.2022.1618.

Khan, Z., & Vorley, T. 2017. Big data text analytics: An enabler of knowledge management. Journal of Knowledge Management, 21(1): 18–34.

Kimberly, J. R., & Evanisko, M. J. 1981. Organizational Innovation: The Influence of Individual, Organizational, and Contextual Factors on Hospital Adoption of Technological and Administrative Innovations. Academy of Management Journal, 24(4): 689–713.

McAfee, A., Brynjolfsson, E., Davenport, T. H., Patil, D. J., & Barton, D. 2012. Big data: The management revolution. Harvard Business Review, 90(10): 60–68.

Suominen, A., Toivanen, H., & Seppänen, M. 2017. Firms’ knowledge profiles: Mapping patent data with unsupervised learning. Technological Forecasting and Social Change, 115: 131–142.

Tödtling, F., & Trippl, M. 2005. One size fits all?: Towards a differentiated regional innovation policy approach. Research Policy, 34(8): 1203–1219.

Tushman, M. L., & O'Reilly, C. A. 1996. Ambidextrous Organizations: Managing Evolutionary and Revolutionary Change. California Management Review, 38(4): 8–29.

Vakili, K., & Kaplan, S. 2021. Organizing for innovation: A contingency view on innovative team configuration. Strategic Management Journal, 42(6): 1159–1183.

Verhoeven, D., Bakker, J., & Veugelers, R. 2016. Measuring technological novelty with patent-based indicators. Research Policy, 45(3): 707–723.

Veugelers, R., & Wang, J. 2019. Scientific novelty and technological impact. Research Policy, 48(6): 1362–1372.

INTRODUCCIÓN

La innovación -tecnológica o creativa- es el concepto más estudiado en los estudios organizativos, dada su colosal implicación en la ventaja competitiva y el rendimiento de una empresa en un entorno cada vez más cambiante (Dess & Picken, 2000; Tushman & O'Reilly, 1996). No se puede exagerar la importancia de la innovación para impulsar el crecimiento, la ventaja competitiva y el éxito general de las organizaciones. Una búsqueda de "innovación organizativa" en Google Scholar arroja miles de artículos que abordan diversas dimensiones de la innovación y la creatividad organizativas, lo que indica su importancia y complejidad en el discurso académico.

Dado su impacto multidisciplinar, el tema de la innovación se ha estudiado en diversas ramas de la ciencia, la ingeniería, las humanidades, las artes y las ciencias sociales, lo que lo convierte en un concepto altamente multidimensional con diversas connotaciones. Se estudia en distintos niveles de análisis, como la persona, el grupo, la organización, la industria y la economía. Godin (2006) identifica más de diez conceptualizaciones relacionadas de la innovación utilizadas históricamente para retratar la innovación a lo largo del tiempo. Esta tesis explora las características internas y externas que ayudan a sintetizar, compartir y dar forma a la innovación organizativa. La innovación no es un proceso único y está determinada por las características únicas del entorno en el que se produce (Tödtling y Trippl, 2005). Por lo tanto, el estudio de los factores contextuales en la investigación de la innovación es fundamental para comprender el entorno complejo y dinámico en el que se produce la innovación (Kimberly y Evanisko, 1981). Los factores contextuales se refieren a las condiciones externas y las circunstancias internas que influyen en el proceso de innovación, incluidos los elementos económicos, institucionales y culturales. Por decirlo de forma poética, el contexto es para la creatividad y la innovación lo que la tierra es para la semilla. Así pues, estudiar la innovación teniendo en cuenta la riqueza contextual puede ayudar a comprender mejor los procesos de innovación y a generar conocimientos teóricos. Este conocimiento es crucial para fundamentar el desarrollo de estrategias y políticas que apoyen y promuevan la innovación.

Los recientes avances en las tecnologías informáticas, como el aprendizaje automático y el procesamiento del lenguaje natural (PLN), han hecho posibles nuevos métodos para comprender y medir la innovación. En poco tiempo, ha surgido un importante corpus de investigación sobre la exploración de los significados integrados en los artefactos de innovación organizativa, como patentes, publicaciones científicas y atributos de productos (Becker, Rullani y Zirpoli, 2021; Bellstam, Bhagat y Cookson, 2021; Boudreau, Guinan, Lakhani y Riedl, 2016; Goldberg, Hannan y Kovács, 2016; Hannigan et al, 2019; Vakili & Kaplan, 2021; Verhoeven, Bakker, & Veugelers, 2016; Veugelers & Wang, 2019). La captura de patrones lingüísticos sutiles y la mejora de las representaciones de constructos a través de herramientas informáticas de procesamiento del lenguaje han amplificado y agudizado el giro lingüístico en la investigación de la innovación (Alvesson & Kärreman, 2000), permitiendo la exploración de los fundamentos cognitivos de la innovación (DiMaggio, 2015). Además, el salto exponencial en el acceso a la información y las capacidades de procesamiento que impulsan la revolución de los macrodatos han convertido la información en un bien de consumo (Khan y Vorley, 2017; McAfee, Brynjolfsson, Davenport, Patil y Barton, 2012). Por ejemplo, ahora es posible agregar y sintetizar los mapas de conocimiento inherentes a la investigación y el desarrollo de productos (Suominen, Toivanen y Seppänen, 2017). Estos avances afectan a todos los aspectos del proceso de investigación e innovación, desde la generación de ideas hasta la difusión de la innovación, y tienen un efecto vigorizante en los estudios organizativos y la teoría de la innovación. Esta tesis apoya la opinión de que la teoría de la innovación puede reforzarse explorando los sutiles mecanismos de aprendizaje organizativo e innovación que posibilita la mayor capacidad para analizar patrones tonales de datos de grano fino (Joseph, Rhee y Wilson, 2022) e intenta aprovechar estos avances para estudiar tres fenómenos interesantes. Esta tesis, cuyo objetivo es contribuir a la teoría de la innovación, presenta tres estudios empíricos para comprender mejor los factores contextuales que afectan a la producción, el intercambio y la síntesis de la innovación. Los tres estudios se sitúan en contextos creativos y de alta tecnología y aprovechan medidas novedosas basadas en textos y avances metodológicos para generar nuevos conocimientos que sirvan de base a la teoría de la innovación.

Los tres estudios empíricos se centran en la comprensión de la producción, el intercambio y la síntesis de la innovación. Estos estudios se sitúan en contextos creativos y de alta tecnología y aprovechan las medidas basadas en textos y los avances metodológicos para generar nuevos conocimientos que sirvan de base a la teoría de la innovación. En el primer capítulo se diseña una metodología para sintetizar y combinar conocimientos procedentes de grandes corpus de patentes y textos de publicaciones académicas con el objetivo de evaluar el valor de la tecnología patentada. El segundo capítulo explora el impacto de la competencia percibida en la difusión de los resultados de la innovación o la I+D en una plataforma reguladora, incluidas las variaciones tonales de la información. El tercer capítulo se sitúa en el contexto único y rico de la industria cinematográfica Bollywood-Hindi, dominada por productores dinásticos. El estudio desarrolla la teoría y pone a prueba la relación entre el arraigo dinástico de los productores y las producciones creativas.

El primer capítulo, "Uncovering The Performance Targets Of Patented Technology Using A Blended Approach Of Topic Modeling And K-Nearest Neighbor Algorithm" (Descubrimiento de los objetivos de rendimiento de la tecnología patentada mediante un enfoque combinado de modelado temático y algoritmo K-Nearest Neighbor), diseña un enfoque combinado que asigna los conocimientos divulgados en publicaciones científicas a invenciones patentadas para descubrir objetivos de rendimiento. Aplicamos nuestra técnica a la investigación sobre almacenamiento de energía eléctrica (EES) para vehículos eléctricos y mostramos cómo puede revelar los objetivos de rendimiento que no se indican explícitamente en el texto de la patente. Nuestro algoritmo combinado obtiene resultados significativamente mejores que el modelado temático de patentes por sí solo a la hora de evaluar la similitud temática. Nuestro estudio pone de relieve la utilidad de combinar fuentes de información alternativas con diferentes niveles de divulgación para descubrir los objetivos de rendimiento de la tecnología patentada, especialmente en tecnologías complejas basadas en componentes.

El segundo capítulo, titulado "Regulatory disclosures and competitive signals: Competition affecting speed and tenor of clinical trial results", sostiene que los mandatos normativos de divulgación que obligan a las empresas a compartir información pueden exacerbar las presiones competitivas. Sin embargo, la bibliografía existente considera el cumplimiento de la normativa como una carga para las empresas que la acatan a regañadientes, pasando por alto la dinámica y las señales competitivas. Combinando los conocimientos de la literatura sobre divulgación de información y la teoría de la señalización competitiva, proponemos que cuando los mandatos de divulgación de información enfrentan a las empresas entre sí, éstas responden a las mayores presiones competitivas acelerando el intercambio de información y adoptando un tenor más positivo en sus divulgaciones. Además, las empresas con reputación de introducir productos novedosos muestran más urgencia y fervor por señalar a sus competidores y reducir la incertidumbre sobre el desarrollo de sus productos. Probamos nuestras hipótesis en el ámbito farmacéutico, aprovechando un conjunto de datos de múltiples fuentes que comprende ensayos clínicos completados entre 2008 y 2019 y registrados en ClinicalTrials.gov bajo el ámbito de la regulación FDAAA 801, información sobre empresas patrocinadoras de Compustat y el Libro Naranja de la FDA, y las publicaciones vinculadas a los ensayos extraídas de PubMed.

El tercer capítulo, "La dualidad dinástica: Examining the sociocultural capital and nepotism discount of Bollywood dynasts", explora la atipicidad del producto de las organizaciones familiares en un contexto cultural creativo. Las productoras familiares son la forma organizativa más extendida en todo el mundo. Les mueve el motivo dinástico: perpetuar su riqueza y estatus a lo largo de generaciones para que los descendientes puedan disfrutar de los frutos del trabajo de sus antepasados. Sin embargo, hay opiniones encontradas sobre si los descendientes de estos productores familiares son innovadores o se ven limitados por el legado de sus antepasados. Por un lado, los descendientes pueden heredar el estatus y los conocimientos tácitos de sus antepasados. Así, disfrutan de las ventajas de acceder a recursos codiciados y de agruparlos, lo que les permite buscar la innovación a través de la tradición. Por otro lado, los descendientes pueden verse limitados por el escepticismo ante el nepotismo y la falta de voluntad de las familias controladoras para profesionalizarse. En este estudio, intentamos comprender el papel de la dinastía en los productores familiares para innovar. Demostramos que las raíces dinásticas de los productores de Bollywood (industria cinematográfica hindú) pueden dar lugar a una curva de innovación en forma de U invertida, en la que el capital sociocultural y el descuento por nepotismo pueden tanto obstaculizar como permitir la innovación.

Los tres capítulos ofrecen perspectivas nuevas y centradas en los elementos esenciales de la innovación, a saber, la síntesis, la divulgación y la producción dentro de los contextos estudiados. Los estudios pretenden construir y poner a prueba teorías sobre la combinación de conocimientos de fuentes académicas e industriales, el impacto de la competencia en la innovación compartida y el papel facilitador o limitador de las conexiones familiares en la creatividad y las innovaciones.

En conclusión, esta tesis se suma a la literatura sobre innovación organizativa, concretamente a la investigación que aborda la producción, el intercambio y la síntesis de innovaciones. Los estudios presentados aportan valiosas ideas a los estudiosos y profesionales analizados en detalle en cada capítulo y sugieren vías para futuras investigaciones en las corrientes de innovación estudiadas.

REFERENCES

Alvesson, M., & Kärreman, D. 2000. Taking the Linguistic Turn in Organizational Research: Challenges, Responses, Consequences. The Journal of Applied Behavioral Science, 36(2): 136–158.

Becker, M. C., Rullani, F., & Zirpoli, F. 2021. The role of digital artefacts in early stages of distributed innovation processes. Research Policy, 50(10): 104349.

Bellstam, G., Bhagat, S., & Cookson, J. A. 2021. A text-based analysis of corporate innovation. Management Science, 67(7): 4004–4031.

Boudreau, K. J., Guinan, E. C., Lakhani, K. R., & Riedl, C. 2016. Looking Across and Looking Beyond the Knowledge Frontier: Intellectual Distance, Novelty, and Resource Allocation in Science. Management Science, 62(10): 2765–2783.

Dess, G. G., & Picken, J. C. 2000. Changing roles: Leadership in the 21st century. Organizational Dynamics, 28(3): 18–34.

DiMaggio, P. 2015. Adapting computational text analysis to social science (and vice versa). Big Data & Society, 2(2): 2053951715602908.

Godin, B. 2006. The Linear Model of Innovation: The Historical Construction of an Analytical Framework. Science, Technology, & Human Values, 31(6): 639–667.

Goldberg, A., Hannan, M. T., & Kovács, B. 2016. What Does It Mean to Span Cultural Boundaries? Variety and Atypicality in Cultural Consumption. American Sociological Review, 81(2): 215–241.

Hannigan, T. R., Haans, R. F., Vakili, K., Tchalian, H., Glaser, V. L., et al. 2019. Topic modeling in management research: Rendering new theory from textual data. Academy of Management Annals, 13(2): 586–632.

Joseph, J., Rhee, L., & Wilson, A. J. 2022. Corporate Hierarchy and Organizational Learning: Member Turnover, Code Change, and Innovation in the Multiunit Firm. Organization Science. https://doi.org/10.1287/orsc.2022.1618.

Khan, Z., & Vorley, T. 2017. Big data text analytics: An enabler of knowledge management. Journal of Knowledge Management, 21(1): 18–34.

Kimberly, J. R., & Evanisko, M. J. 1981. Organizational Innovation: The Influence of Individual, Organizational, and Contextual Factors on Hospital Adoption of Technological and Administrative Innovations. Academy of Management Journal, 24(4): 689–713.

McAfee, A., Brynjolfsson, E., Davenport, T. H., Patil, D. J., & Barton, D. 2012. Big data: The management revolution. Harvard Business Review, 90(10): 60–68.

Suominen, A., Toivanen, H., & Seppänen, M. 2017. Firms' knowledge profiles: Mapping patent data with unsupervised learning. Technological Forecasting and Social Change, 115: 131–142.

Tödtling, F., & Trippl, M. 2005. One size fits all?: Towards a differentiated regional innovation policy approach. Research Policy, 34(8): 1203–1219.

Tushman, M. L., & O'Reilly, C. A. 1996. Ambidextrous Organizations: Managing Evolutionary and Revolutionary Change. California Management Review, 38(4): 8–29.

Vakili, K., & Kaplan, S. 2021. Organizing for innovation: A contingency view on innovative team configuration. Strategic Management Journal, 42(6): 1159–1183.

Verhoeven, D., Bakker, J., & Veugelers, R. 2016. Measuring technological novelty with patent-based indicators. Research Policy, 45(3): 707–723.

Veugelers, R., & Wang, J. 2019. Scientific novelty and technological impact. Research Policy, 48(6): 1362–1372.

Chapter 1: Uncovering the Performance Targets of Patented Technology Using a Blended Approach of Topic Modeling and k-Nearest Neighbor Algorithm.

ABSTRACT

Patents are important resources for firms as a source of competitive advantage, but assessing their value is difficult. Patent statistics are commonly used to measure a firm's innovation activities and performance, but they only explain a small portion of a patent's economic value. Additionally, patents rarely disclose technology performance targets, making it challenging to understand their value. In this study, we propose a blended approach that maps the knowledge disclosed in scientific publications to patented inventions to uncover performance targets. We apply our technique to the Electric Energy Storage (EES) research for electric vehicles and show how it can reveal the performance targets not explicitly stated in the patent's text. Our blended algorithm performs significantly better than topic modeling on patents alone in assessing thematic similarity. Our study highlights the usefulness of blending alternate information sources with different disclosure levels to uncover performance targets of patented technology, particularly in complex component-based technologies.

INTRODUCTION

Patents are widely considered valuable firm resources (Markman, Espina, & Phan, 2004) and are strategically used by firms to gain a competitive advantage (Barney, 1991; Peteraf, 1993). Patent statistics as proxy measures of a Firm's innovation activities and performance have been central to empirical research (Hall, Jaffe, & Trajtenberg, 2005). While patent analysis is common among R&D engineers, academics, and technology policymakers to estimate trends and competition among companies and uncover opportunities, assessing value of a patent is difficult. A patent's value depends on the specific bundle of a legal right, technology quality, and intended use (Reitzig, 2003; Somaya, 2012). The empirical evidence suggests that forward citations and other commonly used patent-based indicators explain less than 5% of the reported economic value of patents (Gambardella et al., 2008). More recently, Sampat & Williams (2019) found patents to have no significant effect on follow-on innovations such as scientific research and product development. These findings highlight the considerable uncertainty around the determinants of patent value and the usefulness of knowledge disclosures in patents.

Patents are designed to appropriate returns from innovations through time-limited exclusive rights and do not necessarily require disclosure of technology performance targets and assessments. Firms are not legally obligated or strategically motivated to disclose such information in their patent applications. They may also strategically omit intended uses of inventions and performance targets from patent texts to protect knowledge spillovers and patent potential substitute and follow-on technologies. Additionally, patents may be used preemptively for strategic purposes with little intention to commercialize the invention. It is particularly complicated to make out performance targets in complex component-based technologies such as semiconductors and electrochemistry, where the value of the technology is fragmented across the patent portfolio instead of a single patent. The lack of knowledge about the performance targets could bias our understanding of the search problem (Jeppesen & Lakhani, 2010), the economic valuation of patents, and, more generally, of targets in technology-lead acquisitions. Hence, it is valuable to ask and investigate: *"How could we uncover performance targets of patented technology?"*

In this study, we leverage differences between R&D activity entailing patents and academic publications, which affect their knowledge disclosures. While patents aim to define the boundaries of IP, tailored to protect and clarify claims on novel processes and products (Lanjouw & Schankerman, 2001). Academic publications prioritize discovery, provide a channel to share knowledge, and foster new scientific contributions. The academic research documented in published articles can be instrumental in understanding hidden paths in corporate (i.e., run by firms) research because institutional norms and strategic motivations surrounding academic publications focus equally on the description of the R&D activity and outcome assessment necessary to formulate meaningful contribution and impact (Dosi, 1982; Rosenberg, 2010; Arora & Gambardella, 1994). It is not trivial to make "secret" knowledge emerge from textual data; however, comparing alternate sources of information with different levels of the disclosure can help this objective (Salandra, 2018). Our devised technique maps the knowledge disclosed in academic publications to patented inventions and helps reveal performance targets.

Our empirical context is Electric Energy Storage (EES) research for electric vehicles. Electric mobility is an important contributor to sustainability, and EES innovations are considered the drivers of the growth of electric vehicles. However, It is a complex system with the possibility of various component and architectural innovations impacting various performance targets. An EES device, or battery pack, is composed of battery modules in combination with control and protection systems, including a cooling system and management system. A battery module is further an assembly of battery cells arranged in protective frames. Electric mobility companies compete in R&D across different technological trajectories, focusing on improving one or more of the various performance targets, such as energy density, charging speed, range, cycle life, heating, corrosion, and cost, that are affected individually and in combination by each EES component (Masias et al., 2021). Therefore, it is an ideal illustrative setting to test our technique as it provides variation in technology and performance targets. We collected all U.S. Patent and Trademark Office (USPTO) patents referring to EES technology granted to firms and academic publications in EES technology during 2001–2017. We devise a blended approach to reveal the performance targets not explicitly stated in the patent's text by reading it through an academic lens created by connecting the patents with *semantically similar* publications. The technique leverages and combines *topic modeling* (Blei, 2012; Blei et al., 2003) *and k-nearest neighbor* (kNN) algorithm (Cover & Hart, 1967; Deng et al., 2016).

To illustrate the validity and value of our *blended algorithm*, we conduct an experiment where we compare the performance of our blended algorithm in assessing the thematic similarity of patents by comparing it with similarity assessments made by knowledgeable and qualified human subjects. Our blended approach performs significantly better than the topic modeling on patents alone. We also conduct seven open-ended interviews with battery engineers and industry experts discussing our results to generate insights about mechanisms and assumptions.

The paper makes the following contributions. We present a novel technique to uncover additional information potentially hidden by companies in their patents, which will help clarify the view on their innovation focus and performance targets. Technology quality and its performance targets are the immanent characteristics of a patent-protected invention and are crucial for the firm's current and future product developments affecting the targeted markets and competition (Nordhaus, 1972). Reading patents and their potentially hidden knowledge could increase information transparency on the market valuation of patents and R&D investment decisions. Our algorithm could serve as decision support in due-diligence processes that are time-consuming and capital-intensive when only performed by human actors. The traditional approaches of patent value or quality assessments requiring both objective and subjective assessments demand high costs, time, and manpower. A scalable method that can shorten the time required to determine the quality of patents concerning their potential use can be very useful (Trappey et al., 2012). Combining our algorithm with human experts at due-diligence processes expedites such processes and increases transparency. Furthermore, such a combination could help overcome potential barriers posed by the limited human knowledge spectrum, allowing human actors to make more informed decisions in due-diligence processes that might not fall into their field of expertise.

The paper is organized as follows: we first provide the theoretical background that relates to our contribution and justifies the underlying assumptions of our approach. We then detail the textual data obtained from the USPTO and top peer-reviewed journals, followed by an explanation of our algorithm. The results section is complemented by interviews with industry experts and an experiment involving knowledgeable human subjects. Finally, we summarize our findings and propose future pathways for extending the research.

BACKGROUND

Patents and Performance targets

Patents provide insight into innovation activities as a window to the knowledge economy (Jaffe & Trajtenberg, 2002). However, a significant portion remains hidden within organizational boundaries (Nelson, 2016). Patenting is a strategic activity where the strategic motivations of the firms guide knowledge disclosure. This strategic behavior raises the question of the quality and performance of patented technology. As Gittelman (2008, page 21) reflects, "Patents are useful indicators but their utility is limited by the extent of our ignorance of process and behaviour that generate them." Personal value (i.e., the value of the patent protection to the patentee) is largely unobservable and proxied by the firm's patent renewal decision (Schankerman & Pakes, 1986; Bessen, 2008). Researchers have struggled with devising better methods or proxies to measure the value of patents and even sought subjective assessments from patentees, experts, or inventors (Gambardella et al., 2008).

Patents are crafted as legal instruments to appropriate the returns from innovations through time-restricted exclusive rights. A patent grants the right to exclude others from using the invention within the delineated boundaries of the verbal claims in the patent for a limited period. Patent examiners apply criteria to ensure the scope of the patent accurately reflects the technological advance described in it. However, due to time constraints and limited access to the prior art, patents are inherently uncertain and imperfect property rights. These uncertainties may only be resolved years after the patent is granted through patent litigation, making the typical patent right quite "fuzzy" in its validity and scope. While the patent law requires the subject matter to be helpful, novel, and non-obviousness, It does not necessarily require information on technology performance assessments. The sufficiency of disclosure or enablement clause in patent law requires the claimed invention to be disclosed in enough detail in a patent application for a person knowledgeable in the field to be able to implement it[1]. Therefore, a patentee firm is neither legally obliged nor strategically driven to disclose its patented invention's performance targets and assessments (Gittelman, 2008; Alcacer & Gittelman, 2006). Firms also strategically omit the intended use of the invention and its performance targets from the patent text so that "firms may patent potential substitute and follow-on technologies themselves (before their competitors)" (Somaya, 2012: p-1092) and by restricting knowledge spillover to competitors. Alternatively, patents have preemptive strategic use with little interest from firms to commercialize these inventions.

This problem multiplies in complex, component-based products where value accrues to a portfolio of patents that can be traded or assembled on component-based products, further complicating the problem of the valuation of a single patent. For instance, a battery system is an assembly of various components, each simultaneously affecting various performance parameters, such as lifespan, safety, performance across

[1] See enablement requirement as stated by the United States Patent and Trademark Office (USPTO) and World Intellectual Property Organization (WIPO) @ https://www.uspto.gov/web/offices/pac/mpep/s2164.html and https://www.wipo.int/patents/en/faq_patents.html respectively.

temperatures, charging time, and range (Dinger et al., 2010). These performance concerns require improving one or more technical parameters, such as energy density, power density, specific power, specific energy, cycle life, and electrochemical impedance. The parameters, in turn, depend on the proprietary electrochemistry of the battery cell, the most fundamental component of an EES device, and the overall design or architecture of the EES. It is a complex system to ascertain the value of each technological tweak. Any meaningful assessment requires painstaking testing efforts and is often absent from the patent text.

Patents, Academic Publications, and Performance Targets

Patents and academic publications serve distinct purposes for technology development and dissemination (Nelson, 2016). While knowledge sharing is both a norm and an institutional imperative (or "institutional logic") in publications, commercial engagement, such as patenting, dampens sharing by limiting sharing (Campbell, Weissman, Causino, & Blumenthal, 2000). Patents, as legal documents, are primarily used to protect the inventor's intellectual property by establishing exclusive rights to manufacture, use, and sell the described technology. In contrast, Academic publications showcase the performance and functionality of new technology and establish its superiority to attract interest from potential customers, partners, and investors. Academic articles are typically peer-reviewed and undergo rigorous scrutiny to ensure that the research presented is of high quality and contributes to the existing body of knowledge. It makes them an excellent instrument for marketing the technology. Scientific publications contribute to establishing research credibility and signal the potential commercial value of the technology (Rotolo et al., 2022). Therefore, while patents and publications play essential roles in technology development, they serve distinct purposes and entail subtle differences in disclosed knowledge.

Academic publications prioritize discovery, provide a channel to share knowledge, and foster new scientific contributions. These publications are primarily used to disseminate the technology gains by focusing on performance superiority and assessments. Patents, instead, aim to define the boundaries of IP, tailored to protect and clarify claims on novel processes and products (Lanjouw & Schankerman, 2001). The publications are peer-reviewed by experts in their respective fields with a superior technological understanding and require clear contribution statements focusing more on the "So what" rather than novelty. The contribution and impact are more crucial in scientific communication than novelty. In contrast, Patent examiners are public officers who review patent claims. With a certain legal and technical qualification, these officers are employed to assess patent applications in accordance with legal rules and recommended novelty (Gans et al., 2008). Therefore, these two document types, i.e., patents and academic articles, differ not only in their purpose but also in terms of what knowledge gets disclosed and how. Given the difference in perspectives, norms, and sense of contribution - Patents and scientific publications often provide complementary facets of innovations. Leveraging these different and complementary characteristics of patents and publications, we devise a blended approach to shed more light on the undisclosed parts of firm innovation, specifically -performance targets.

DATA AND SETTING

Electric Energy Storage (EES) Technology For Electric Vehicles

Power generation in electric vehicles relies on two main technological paradigms: fuel cells that generate electricity from hydrogen or EES devices comprised of batteries (Pollet et al., 2012). These configurations rely on substantially different underlying scientific principles (Cano et al., 2018). To ensure that we investigate comparable trajectories, we focus only on the battery paradigm, which has dominated the electric vehicles market thus far.

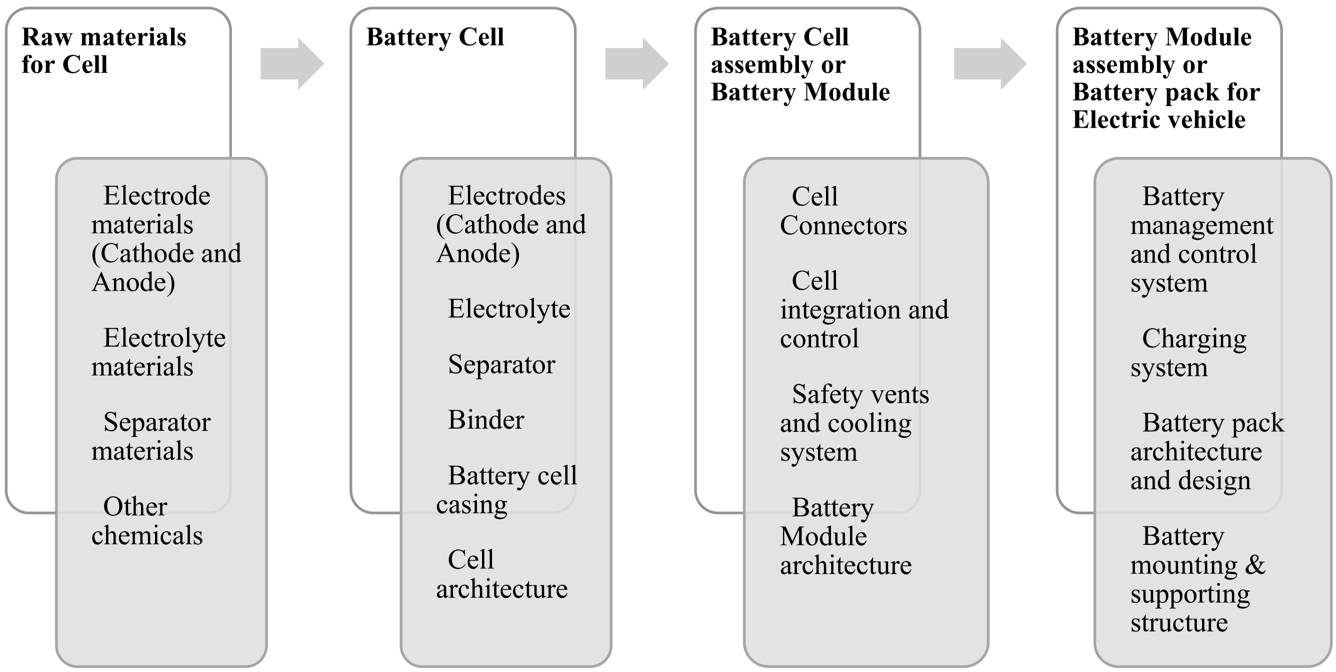
Advances in EES technology have increased the efficiency of electric vehicles, which in turn are disrupting the automobile industry (Dinger et al., 2010). Given its sustainability impact, EES technology is a high priority for both public research and for-profit companies. EES will pave the way toward full electric mobility, reducing societal carbon footprint and increasing sustainability (Valogianni et al., 2020). For firms, the technology will secure a competitive advantage against conventional internal combustion engine cars (Lim et al., 2015) and bring high stock returns. In this high-technology industry, with strong links to the academic fields of electrochemistry and materials science (Pollet et al., 2012), academic researchers can obtain substantial government funding; thus, electric mobility research is vigorously pursued by research centers and universities[2].

Each EES or battery pack contains various modular components and is designed to meet a particular vehicle's specifications. For instance, a battery system is an assembly of various components, each simultaneously affecting various performance parameters (Dinger et al., 2010). Along with cost and size criteria, some critical challenges for EES include lifespan, safety, performance across temperature ranges, charging time, and charge behavior (Lim et al., 2015; Rauh, et al., 2015; Ma et al., 2022). These performance concerns require improving one or more technical parameters, such as energy density, power density,

[2] See https://www.energy.gov/technologytransitions/downloads/august-2018-spotlight-solving-challenges-energy-storage [Accessed:27-November-2020].

specific power, specific energy, cycle life, and electrochemical impedance. The parameters, in turn, depend on the proprietary electrochemistry of the battery cell, the most fundamental component of an EES device, and the overall design or architecture of the EES. Figure-1 provides an illustration of the architecture and value chain of an EES device for electric vehicles, with its main components and basic design hierarchy.

Figure 1. An illustration of the Basic architecture and value-chain of EES or battery Pack for Electric vehicles in terms of its composition.



(Golembiewski, vom Stein, Sick, & Wiemhöfer, 2015)

DATA DESCRIPTION

We collected two document sets: patents granted by the USPTO and academic articles published in high-impact journals. For the patent document set, we retrieved from USPTO the entire corpus of granted patents in the technological field of EES devices for electric vehicles from 2001 to 2017. These patents are granted under the Cooperative Patent Classification (CPC) class "H01M" and include ("electric vehicle" or "electric car") and ("battery" or "batteries") anywhere in their entire text. In total, we retrieved 9985 patents granted to private, for-profit firms between 2001 and 2017.

For the publications dataset, we searched Thomson Reuters Web of Science Core Collections for articles published between 2001 and 2017 dealing with electric vehicle battery technology. Specifically, We used advanced topic search using the following query TS=(electric* AND (car OR vehicle* OR mobility) AND (battery OR batteries)), which provided all the articles with defined terms in the Title, Abstract, or Keywords record fields. We restricted our search to articles in the English language and excluded conference proceedings. This gave us 5090 articles published in 865 unique journals. To ensure we capture high-quality scientific research, we selected journals occupying JIF quartile Q1 and Q2 as per Clarivate's Journal Citation Report 2020. Finally, We are left with 3437 articles dealing with electric vehicle battery technology published in 330 high-impact and most prestigious journals between 2001 and 2017 (both years included).

The USPTO patents are uniformly structured, and information about the innovation appears under four main headings: "Title," "Abstract," "Claims," and "Description." Each patent document has an identifying title, an abstract that briefly describes the innovation, a description section that identifies its application space, and a claims section that lists the proprietary innovation claims secured by the patent. We include the entire text available in the patent data set. Academic publications are semi-structured documents; they do not adhere to universal requirements or norms. The structure of documents and the text largely depends on the research problem, methods used, findings, and journal guidelines (Perneger & Hudelson, 2004). However, they all have titles, abstracts, and keywords conveying comparatively uniform information and summarizing the major knowledge gains, so we use these three sections combined as appropriate text input from the article data set.

METHODOLOGY

Blended Algorithm

To understand firms' innovation objectives and performance targets using these big data sets, we design an algorithm that has *topic modeling* (Blei, 2012), and specifically Latent Dirichlet Allocation (LDA) (Blei et al., 2003), at its core. The presented algorithm is comprised by two main steps: 1) topic modeling to elicit thematic topics from the patent and article data sets, 2) kNN semantic similarity computation between patents and articles to establish the semantic proximity of each patent with articles to identify the latent article knowledge theme in each Patent. Each of these steps are described in detail below. As a first step, the presented algorithm applies topic modeling on industrial firm patents, allowing us to understand how firms approach EES research, as captured in their patent texts. This first step aggregates and summarizes the large volumes of firm patents into themes (topics) understandable by humans, allowing human experts to have a clearer overview of innovation positions of firms, as reflected in their patents.

Then, we read the firm patents differently (i.e., through an academic lens) by connecting the patents into consideration with *semantically similar* scientific articles published during the same period. This becomes possible not only because technological advances in EES are reflected in scientific articles about electrochemistry and materials science (Pollet et al., 2012), but also because academic inventors are less concerned about imitation threats, making them more open toward disclosing knowledge in public texts. To obtain this semantic similarity, we create an *exogenous vocabulary map* using academic publications in top journals by academic authors without firm affiliation related to EES technologies during the same period (2001–2017). More specifically, our algorithm calculates the textual and semantic similarity of each patent to all scientific publications in the input data set, using the k-nearest neighbor (kNN) algorithm (Cover & Hart, 1967; Deng et al., 2016). Using the semantically most similar article of each patent as the closest proxy, we can re-classify the patents using the topic taxonomy derived from the corpus of academic articles. In other words, we are reading each patent under an academic lens with the help of the semantically similar publications. This help identify components that are not explicitly mentioned or immediately apparent in firm patents; however, they are more clearly elicited from their academic proxies. As a result, our algorithm links the invention with the academic knowledge themes, providing a view that is helpful in linking firms with other firms targeting similar usage and performance targets. These targets are often generally suppressed in patent texts. This approach offers a comprehensive and effective method for identifying firms' innovation objectives and performance targets

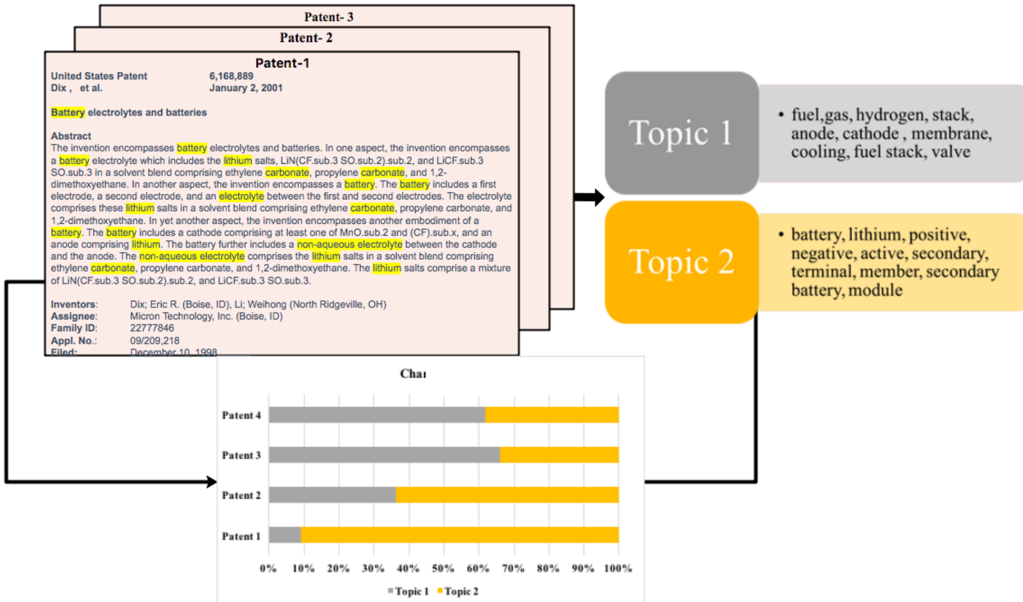
Topic Modeling (LDA)

As highlighted by Aras et al. (2014), the patent text is rather complex to parse and process correctly. Hence, a certain pre-processing is required before using the text. We follow a series of procedures for cleaning the textual data used in the machine learning literature (Mikolov et al., 2013; Aras et al., 2014). First, we remove punctuation, words that carry low informational value (such as words “end”, “the”, etc.). Furthermore, an important element in this step is that we also account for two-word phrases (such as “New York”), which might have a different meaning than the two words separately (e.g., “new” and “York”).

To identify the topics discussed in each patent or article in the cleaned data sets, we implement Latent Dirichlet Allocation (LDA) (Blei et al., 2003). LDA is an unsupervised topic modeling algorithm that identifies key topics discussed in texts, in our case, academic and corporate research. LDA offers robust topic modeling that can uncover prominent themes reflected in the text. It has shown superior performance in diverse text mining environments, making it one of the state-of-the-art for topic modeling. LDA assumes documents as mixtures of probability distributions over topics of discussion.

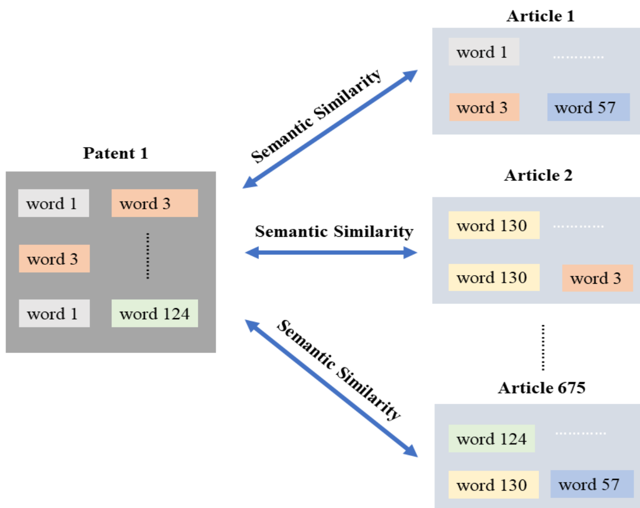
Furthermore, each topic is comprised of probability distributions over words. Hence, each document in our data sets is considered a mixture of a small number of underlying discussion topics that appear with certain probabilities. As a result, taking only the text in documents as input, LDA generates automatic summaries of prominent topics discussed in the text (topics of discussion and their probability of being present in the document in question). The main advantage of LDA against other topic modeling methods is that it can yield human-understandable topics, which is desirable in this particular context. Figure 1 presents an exemplary topic distribution elicitation for Patent-6168889 of our data set. For expositional simplicity, we assume that LDA elicits two topics (Topic 1 and Topic 2) from this patent in this example. As shown in the bottom percentage graph of Figure 2, Patent-6168889 (denoted as Patent-1) has an over 90% representation of terms belonging to Topic 2 in its patent text; hence, it is assigned to Topic 2. A similar rationale is followed for all patents and academic articles in our data sets. The specifics of our LDA algorithm and the selection of the number of topics are presented with results.

Figure 2. Exemplary topic distribution elicitation for Patent-6168889 of our patent data set



kNN- Reclassification of patents based on text similarity with academic articles After eliciting the underlying topics from text datasets of academic articles and patents separately, we apply the k-Nearest Neighbor (kNN) algorithm to calculate the cosine semantic similarity of patents and articles (Altman, 1992). This algorithm is powerful but not too computationally complex for text categorization (Guo et al., 2006), hence, suitable for our approach, where we have large volumes of textual data. The main rationale of this step is that for each patent, which is a combination of words, we find the most semantically similar articles. A schematic overview of the semantic similarity calculation is shown in Figure 3.

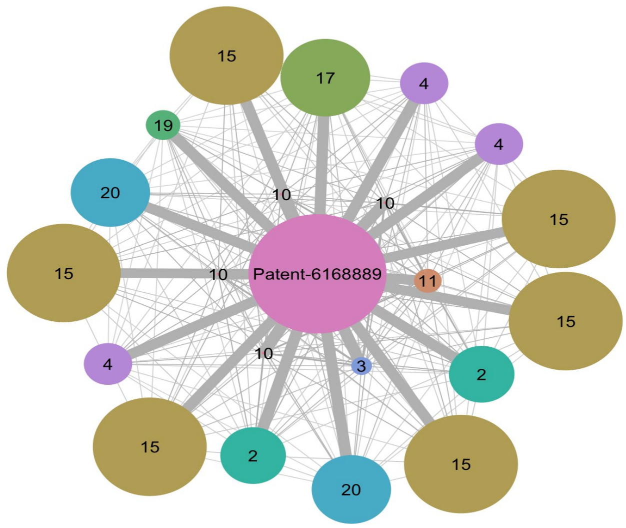
Figure 3. Schematic overview of the semantic similarity calculation between each patent and all available academic articles



Our ultimate goal is to categorize patent documents in the academic article topics. In other words, to get an "academic proxy" for patents. Therefore, the implemented kNN algorithm first, calculates the semantic distance between a patent and each article in our data set. Then it sorts all semantic distances by increasing order and, if *n* indicates the number of topics in LDA, it selects the *n+1*[3] nearest neighbor articles of the patent. We need to have at least *n* nearest neighbors, so that we make sure that all *n* topics have the potential to be represented in the nearest neighbors. As a result, the patent in consideration is labeled with the topic which represents the most frequent theme in its *n+1* semantically most similar articles. Figure 4 illustrates this approach using Patent– 6168889 (Title: Battery electrolytes and batteries) as an example. Let us assume, for demonstration purposes, that LDA had *n=20* topics. This patent was found by LDA on patent data to belong in the topic "Advanced Li ion batteries - Electrolyte composition."

[3] We also perform a robustness check with *n* nearest neighbors in case there are significant changes in the results.

Figure 4. Illustration of kNN classification of a patent based on semantic similarity with academic articles.



Note: The node size denotes the semantic similarity value, with the most similar article having the largest node.

However, when calculating the kNN similarity of this patent's term vector to the academic article vectors, 6 out of 21 nearest neighbors of the patent belong to academic article topic 15. Hence, this patent is considered most semantically similar with academic articles belonging to topic 15, which is the topic with the highest frequency of occurrence in these 21 nearest articles.

RESULTS

LDA: Academic and industrial topics

Selection of number of topics In order to select the number of topics n for LDA of our algorithm, we use a set of criteria proposed in the literature (Griffiths & Steyvers 2004; Cao et al., 2009; Lau et al., 2018; Sbalchiero & Eder, 2020), in which the optimal number of topics is determined based on semantic coherence. To find the optimal topic solution, we calculated and plotted the coherence for various topic solutions up till 40 topics. Figure

5 presents the coherence score plot for Articles dataset comprising of 3437 abstracts and figure 6 presents the coherence score plot for Patents.

Figure 5. Coherence scores plot for various topic solutions for Articles dataset up till 40 topics.

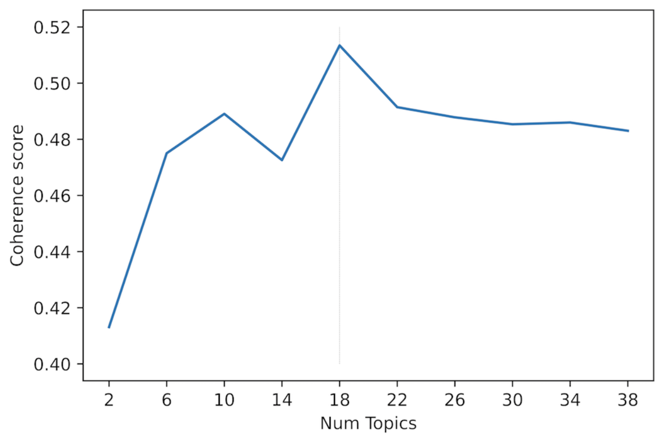
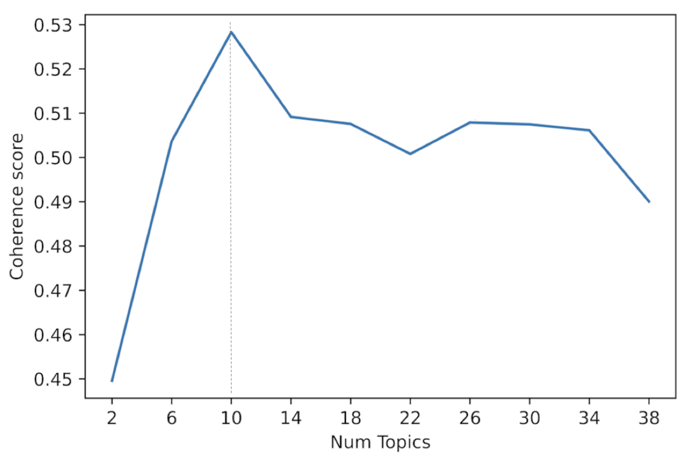


Figure 6. Coherence scores plot for various topic solutions for Patents dataset up till 40 topics.



We find optimal and select 18 topic solution for Articles dataset as per this criteria(with max coherence score of 0.5134) and 10 topic solution for the patents dataset as per this criteria(with max coherence score of 0.5283).

Theme identification and Topic labeling One of the main aims of the LDA algorithm is to extract meaningful topics, which represent the identifiable and unique themes. Hence, topic labeling serves as another test of topic solution quality. The labeling involves identifying the unique theme in each topic and its associated terms. We follow a dual approach to label the identified themes. First, to identify a common unifying theme, we investigate the most frequent terms in each topic (i.e., the words with the highest probability of belonging in a topic) and consider them along with the Titles of most representative documents of each topic (documents with more than 90% probability of belonging to that topic). Following this approach, we are able to identify distinct themes, which correspond to different research topics.

Academic Research Topics In Table 1, we present the 18 topics as identified via the LDA from data set $\boldsymbol{A'}$, and their relative prominence, measured in terms of number of articles belonging in them.

Table 1. 18-topic LDA solution for academic articles A' with identified themes.

Topics/Themes	Top 10 Keywords With Associative Probabilities	Number Of Documents
(1) Electric Vehicles Charging And Driving Route Anaysis	0.044*"electric_vehicle" + 0.025*"time" + 0.017*"result" + 0.016*"datum" + 0.011*"analysis" + 0.010*"base" + 0.010*"network" + 0.009*"level" + 0.009*"driver" + 0.009*"route"	144
(2) High Capacity /Performance Li-Ion Battery Composition	0.073*"high" + 0.036*"performance" + 0.034*"capacity" + 0.022*"lithium_ion" + 0.021*"cycle" + 0.021*"electrode" + 0.017*"composite" + 0.016*"anode" + 0.015*"carbon" + 0.013*"cathode"	323
(3) Li-Ion Battery Capacity Degradation And Aging	0.134*"cell" + 0.069*"battery" + 0.063*"lithium_ion" + 0.030*"capacity" + 0.019*"li_ion" + 0.019*"cycle" + 0.018*"degradation" + 0.017*"age" + 0.015*"rights_reserve" + 0.014*"test"	188
(4) Technology Impact And Emission Analysis Of Electric Vehicles	0.034*"emission" + 0.023*"technology" + 0.021*"vehicle" + 0.019*"car" + 0.019*"life_cycle" + 0.018*"fuel" + 0.018*"analysis" + 0.017*"impact" + 0.015*"electric" + 0.014*"compare"	265
(5) Battery Life And Performance Analysis	0.270*"battery" + 0.034*"charge" + 0.030*"test" + 0.028*"performance" + 0.024*"lead_acid" + 0.023*"life" + 0.018*"discharge" + 0.016*"application" + 0.015*"develop" + 0.014*"elsevi"	122
(6) Electric Vehicle Cost Analysis And Market Potential	0.171*"vehicle" + 0.074*"cost" + 0.057*"plug" + 0.028*"electric" + 0.016*"phev" + 0.016*"range" + 0.014*"fleet" + 0.011*"potential" + 0.010*"market" + 0.010*"phevs"	135
(7) Lithium Ion Battery Electrolyte Composition	0.050*"lithium" + 0.035*"electrolyte" + 0.021*"battery" + 0.020*"base" + 0.015*"air" + 0.014*"process" + 0.014*"mobility" + 0.013*"metal" + 0.011*"ion" + 0.011*"polymer"	191
(8) Battery Voltage Converters And Efficiency	0.044*"converter" + 0.039*"voltage" + 0.026*"propose" + 0.024*"control" + 0.021*"current" + 0.019*"design" + 0.017*"high" + 0.014*"operation" + 0.012*"efficiency" + 0.012*"dc_dc"	264

(9) Hydrogen Based Energy Storage System	0.170*"energy" + 0.027*"storage" + 0.026*"system" + 0.020*"hydrogen" + 0.020*"efficiency" + 0.018*"source" + 0.018*"demand" + 0.017*"power" + 0.017*"base" + 0.014*"supply"	101
(10) Electric Vehicles As Loads In Smart Grids	0.089*"charge" + 0.054*"grid" + 0.025*"electric_vehicle" + 0.023*"evs" + 0.019*"load" + 0.019*"distribution" + 0.015*"impact" + 0.014*"operation" + 0.013*"demand" + 0.012*"service"	312
(11) Battery Thermal Design And Electric Vehicle Performance	0.109*"model" + 0.070*"battery" + 0.037*"temperature" + 0.029*"thermal" + 0.026*"simulation" + 0.023*"design" + 0.015*"performance" + 0.013*"electric_vehicle" + 0.012*"cool" + 0.012*"condition"	206
(12) Hybrid Electric Vehicle Design And Powertrain Efficiency	0.095*"hybrid" + 0.077*"vehicle" + 0.032*"design" + 0.032*"electric" + 0.024*"drive" + 0.023*"efficiency" + 0.023*"engine" + 0.022*"performance" + 0.020*"fuel_cell" + 0.016*"powertrain"	235
(13) Optmized Eneregy Management And Control	0.056*"control" + 0.054*"strategy" + 0.039*"optimization" + 0.034*"base" + 0.031*"optimal" + 0.031*"propose" + 0.027*"energy_management" + 0.019*"problem" + 0.017*"controller" + 0.016*"management"	227
(14) Develoments In Battery Technology And Applications	0.090*"battery" + 0.038*"application" + 0.035*"high" + 0.027*"development" + 0.023*"technology" + 0.021*"electrical" + 0.020*"supercapacitor" + 0.020*"device" + 0.017*"make" + 0.015*"energy_storage"	103
(15) Battery Cathode Performance	0.055*"increase" + 0.037*"show" + 0.037*"result" + 0.035*"study" + 0.034*"battery" + 0.029*"low" + 0.028*"reduce" + 0.026*"large" + 0.024*"due" + 0.022*"improve"	20
(16) State Of Charge Estimations Of Battery	0.075*"battery" + 0.072*"state" + 0.070*"model" + 0.040*"method" + 0.031*"base" + 0.023*"parameter" + 0.022*"estimation" + 0.019*"soc" + 0.018*"charge" + 0.018*"estimate"	279
(17) Electric Vehicle Power Transfer System Management	0.239*"system" + 0.191*"power" + 0.024*"electric_vehicle" + 0.020*"dynamic" + 0.017*"ultracapacitor" + 0.016*"energy_storage" + 0.016*"paper" + 0.013*"present" + 0.011*"management" + 0.010*"base"	100
(18) Electrode Material For High Energy Density Battery	0.044*"material" + 0.025*"high" + 0.025*"electrode" + 0.019*"li_ion" + 0.017*"structure" + 0.017*"electrochemical" + 0.015*"surface" + 0.014*"cathode_material" + 0.014*"phase" + 0.013*"show"	222

As these results show, academic research (data set $\boldsymbol{A'}$) is mainly concerned with the performance aspects of battery storage components and devices (e.g., "monitoring", "temperature performance", "aging", "charging time"). These findings support over view that peer-reviewed academic articles focus on technology performance assessments and issues (Hessels & van Lente, 2008).

Industrial Research Topics LDA uncovers 10 topics from the patent data set $\mathbf{P'}$as shown in Table 2.

Table 2. 10-topic LDA solution for patent data set P′ with identified Themes.

Topics/Themes	Top 10 Keywords With Associative Probabilities	Number Of Documents
(1)Battery Electrode Materials	0.056*"layer" + 0.037*"electrode" + 0.031*"material" + 0.020*"form" + 0.018*"separator" + 0.017*"collector" + 0.017*"battery" + 0.016*"surface" + 0.015*"active" + 0.015*"current"	726
(2)Battery Cathodes And Metals	0.019*"material" + 0.016*"metal" + 0.016*"electrolyte" + 0.015*"cell" + 0.015*"comprise" + 0.011*"claim" + 0.010*"electrochemical" + 0.009*"invention" + 0.008*"alloy" + 0.008*"cathode"	1135
(3)Battery Charging Systems	0.064*"electrode" + 0.042*"plate" + 0.030*"portion" + 0.028*"terminal" + 0.028*"battery" + 0.023*"positive" + 0.022*"negative" + 0.018*"member" + 0.018*"case" + 0.017*"assembly"	1055

(4)Voltage Control Systems	0.040*"voltage" + 0.037*"battery" + 0.036*"power" + 0.031*"charge" + 0.024*"control" + 0.021*"current" + 0.018*"unit" + 0.017*"electric" + 0.015*"device" + 0.014*"state"	1269
(5)Lithium Battery Systems	0.044*"lithium" + 0.037*"material" + 0.034*"electrode" + 0.030*"battery" + 0.023*"active" + 0.019*"positive" + 0.017*"negative" + 0.015*"secondary" + 0.012*"ion" + 0.011*"include"	1569
(6)Battery Packs	0.175*"battery" + 0.114*"cell" + 0.051*"module" + 0.038*"pack" + 0.023*"unit" + 0.019*"plurality" + 0.016*"terminal" + 0.015*"fig" + 0.014*"connect" + 0.013*"include"	675
(7)Battery Structure	0.030*"portion" + 0.027*"battery" + 0.023*"side" + 0.021*"member" + 0.020*"fig" + 0.016*"end" + 0.015*"surface" + 0.013*"form" + 0.013*"plate" + 0.011*"cover"	1320
(8)Fuel Cell Systems	0.101*"fuel" + 0.081*"cell" + 0.037*"gas" + 0.025*"hydrogen" + 0.022*"system" + 0.021*"supply" + 0.020*"stack" + 0.014*"water" + 0.014*"air" + 0.012*"power" + 0.012*"pressure"	998
(9)Battery Heating And Collign Systems	0.052*"cool" + 0.047*"temperature" + 0.034*"air" + 0.026*"heat" + 0.023*"flow" + 0.019*"coolant" + 0.017*"battery" + 0.017*"fluid" + 0.016*"vehicle" + 0.014*"system"	497
(10)Power Storage Devices And Systems	0.036*"system" + 0.029*"power" + 0.026*"device" + 0.023*"vehicle" + 0.023*"energy" + 0.020*"electrical" + 0.019*"battery" + 0.016*"include" + 0.016*"storage" + 0.013*"charge"	741

Corporate research topics do not reveal explicit battery performance metrics. These findings are qualitatively verified by our interviews with electric mobility engineers and innovators, one of which explained that “electric vehicle applications are very heavily concerned with the pack factors like space efficiency, cooling … there is a lot of design effort.”

Interviews with Industry Practitioners

To get a more definite sense of the meaning of these findings, we contacted seven engineers with product R&D experience in EES or the automotive industry. In an open interview setting, involving interviews that lasted 45 minutes on average and were recorded and then transcribed, we began by explaining the research aims and methodology. Then we asked these expert respondents to comment on the results. After contemplating these summaries, the respondents indicated they found the results consistent with their views of research trajectories in EES, calling them “not surprising” in the words of one respondent. They also offered explanations for the prominence of modular innovation in academic research, and architectural innovation in industrial research, namely, differences in the conceptualization of novelty between firms versus academics. One respondent explained, "it's difficult to patent on few component enhancements, because you need really to find the novelty and for the novelty to be distinguished from the others. Novelty is in the combination." The architectural approaches, characterized by innovative combinations of components, make it easier to claim novelty to satisfy patent examiners. An engineer offered an example of trade-offs in architectural design. Noting that inserting more components and cables might increase several performance measures but also would increase the battery's weight and volume (which are critical commercial measures in the automotive industry), he stated that "electric vehicle applications are very heavily concerned about the pack factors like space efficiency, cooling … there is a lot of design effort."

Moreover, the interviews suggest that companies "are talking about the system level; at the vehicle level, they won't talk about or compare individual battery cells. Batteries are energy resources that are utilized in your vehicles, so your energy management strategy is what becomes important, and real competitive advantage would be in how you manage that energy resource and how you manage your other components that are feeding on that energy." Another engineer agreed: "the primary focus of the electric vehicle firms is to improve architecture and assembly design taking basic components as given."

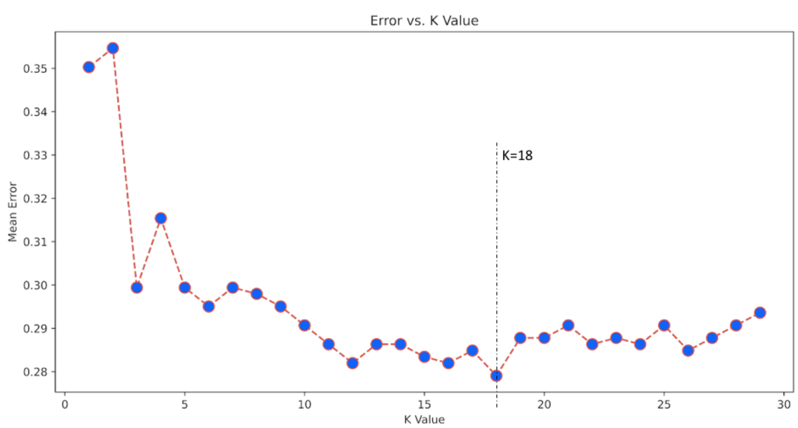
These explanations triggered us to ask whether companies use selective revealing and withholding in writing their patent applications. The interviewees indicated that patent writing "is really a company strategy," and firms have scarce incentives to share their performance aims. Specific applications rarely are inserted in patents, as an enhanced protection measure, which also helps reduce imitability risks across non-protected applications. A respondent illustrated this point with an example: "we have a technology … then someone was copying our technology for other purposes, and we cannot avoid its patents. Hence, when you have an idea, you might not know all the applications, and hence you patent the general technology so, in the future, you can develop more applications." Along similar lines, an interviewee noted that "this is why, in patents, it is very difficult to find solutions to very specific performance topics; this idea of applications is hidden. This is a strategy to block competition, or in the future, you use patents for a negotiation." Another expert recognized, "once you are granted a patent, that information is available to everyone. That information can be used in many different ways, so rather than putting it in a patent, sometimes they will just hold back and continue with developing a product and their research."

Next, we discussed whether interpreting patents according to academic research topics could help reveal some hidden patterns. One respondent showed us how interpreting architectural approaches with a "hidden" map of components and materials (and corresponding performance targets) could help reveal the directions companies were pursuing. Another engineer specified even more clearly that "when electric vehicle producers pick up a cell, they know the performance of that cell, which has already been established. They do the rest of the work taking that as the base." Finally, the respondents confirmed that companies explore different R&D trajectories to try to build their competitive advantages, such that "everybody has a different specification. Battery and electric vehicles together are a selling point. Firms are trying different combinations. Today the focus is on uniqueness and not standardization of battery packs." These arguments also emerge in the popular press, which has reported that "the differing strategies show a division within the auto industry over what is the best path to full electrification"[4].

Mapping Industrial Patents to Academic Research Topics

As a final step of our blended approach, we perform reclassification of patents in the set $\mathbf{P}'$ using the k-Nearest Neighbor (kNN) algorithm based on the cosine text similarity with academic articles $\boldsymbol{A}'$. The algorithm calculates the semantic similarity of each patent with all the articles in the articles dataset A', and selects the k most similar articles as neighbors. To determine the optimal value of k, we plotted the error rate for the articles dataset for values of k ranging from 1 to 30, and found that k=18 provides the best performance. Figure 7 shows the plot of the mean error for the predicted values of the test set for all the K values between 1 and 30.

Figure 7. Plot of the mean error for the predicted values of the test set for all the K values between 1 and 30.



---

[4] See https://www.wsj.com/articles/gm-volkswagen-say-goodbye-to-hybrid-vehicles-11565602200 [Date Accessed: 28-Nov-2019].

Another important consideration is the similarity threshold used for reclassification. To determine this, we randomly selected 100 patents from our sample and extracted all patent citations/references, resulting in 131 valid article-patent pairs. We then downloaded the abstracts for all these and calculated the cosine similarity between the articles and corresponding patent texts. The mean cosine similarity was 0.09645, which we rounded off to 0.1. We chose this mean similarity threshold of 0.1 as sufficient for reclassification.

The classifier was then applied in a loop to classify 9985 patents based on cosine distance with 3437 articles and successfully reclassified 8047 patents with article themes. The most prominent themes that emerged after reclassification are (3) Li-Ion Battery Capacity Degradation And Aging, (11) Battery Thermal Design And Electric Vehicle Performance and (2) High Capacity /Performance Li-Ion Battery Composition.

Experimental Evaluation Of Patent Thematic Similarities With Human Respondents

So far, we have shown the differences in the firm's patent themes using the topic modeling assessment of patent texts and our synthesized technique combining the patents and academic articles information (i.e., Blended algorithm). We construct and launch an experiment with the human respondents to test and validate if *Blended algorithm* captures meaningful thematic similarities compared to the simple topic modeling approach. We aim to measure text similarities perceived by human subjects and use these as a benchmark to analyze and compare topic modeling and *Blended algorithm*.

Experiment Design and Results

Considering the thematic similarity assessments of the patents by topic modeling and *Blended algorithm*, we carefully select two sets of three patents each from our sample for the following experimental manipulations. We are going to manipulate the level of similarity across different patents. In experiment set 1, we randomly choose three patent pairs where two pairs have low *topic modeling similarity*, and one pair has high *topic modeling similarity*. All these three patents of set 1 should have low *Blended algorithm similarity*. In experiment set 2, we choose three patent pairs where two pairs have low *Blended algorithm similarity*, and one pair has high *Blended algorithm similarity*. All these three patents of set 2 have low *topic modeling similarity*. Therefore, we prime respondent with six patents with varying degrees of similarities; respondents are agnostic about this manipulation. The details of the experiment sets and patent pair combinations are in Table 3[5].

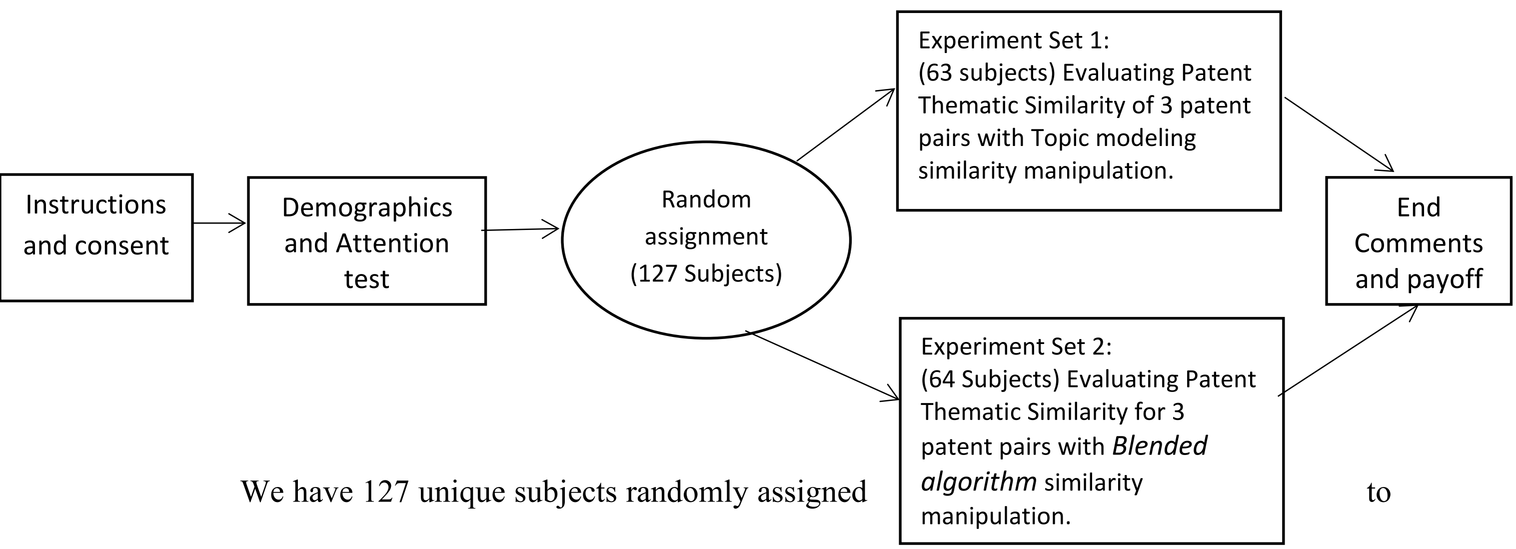
Table 3. Details of the experiment sets and patent pair combinations.

Experiment set	Patent pair	Patent pair titles	Topic Modeling similarity (1:high;0:low)	Blended algorithm similarity (1:high;0:low)
1	1	(System and method for maximizing a battery pack total energy metric) X (Thermal energy transfer system for a power source utilizing both metal-air and non-metal-air battery packs)	1	0
1	2	(System and method for maximizing a battery pack total energy metric) X (High storage capacity, fast kinetics, long cycle-life, hydrogen storage alloys)	0	0
1	3	(Thermal energy transfer system for a power source utilizing both metal-air and non-metal-air battery packs) X (High storage capacity, fast kinetics, long cycle-life, hydrogen storage alloys)	0	0
2	4	(Hybrid battery system) X (Method for controlling charging voltage of 12V auxiliary battery for hybrid vehicle)	0	1
2	5	(Method for controlling charging voltage of 12V auxiliary battery for hybrid vehicle) X (High storage capacity alloys having excellent kinetics and a long cycle life)	0	0
2	6	(Hybrid battery system) X (High storage capacity alloys having excellent kinetics and a long cycle life)	0	0

[5] The Details of Patents selected for the two experiment sets are logged in Appendix that is available on request.

The participants for this study were selected using the Qualtrics XM Survey Platform, satisfying the inclusion criteria of being native English speakers who had attained at least a graduate degree. All participants who completed the experiment were compensated equally. The median completion time was 8 minutes and 50 seconds, and no single nationality comprised a dominant portion of the sample. The Subjects were randomly assigned to two experiment sets (5 being the highest and 1 being the lowest similarity). Figure 8 presents the sequence of events within the experiment[6].

Figure 8. Sequence of events within the experiment.



We have 127 unique subjects randomly assigned to experiment set 1 (63 subjects) and experiment set 2 (64 subjects). In each experiment set, a subject encounters three patent text pairs and is asked to make three similarity assessments for each of the three patent pairs, giving us 381 (i.e., 127*3) responses. We removed the responses that failed the attention test (4 subjects failed attention test, invalidating 12 responses) and one response was without similarity value. Finally, we are left with 368 valid responses. All subjects are highly educated and have college degrees;

[6] The detailed screenshots of the experiment and survey are logged in Appendix that is available on request.

about two third are post-graduates. Almost fifty-five percent identify as male. Table 4 and 5 report the descriptive statistics and correlations.

Table 4. Descriptive Statistics

Variable	Obs	Mean	Std. Dev.	Min	Max
Patent pair similarity assessments					
Perceived Similarity	368	3.48	1.22	1	5
Topic Modeling Similarity	368	0.17	0.37	0	1
Blended algorithm Similarity	368	0.17	0.37	0	1
Gender					
Male	368	0.54	0.5	0	1
Female	368	0.42	0.49	0	1
Non-binary	368	0.02	0.15	0	1
Did not disclose	368	0.02	0.13	0	1
Knowledge about battery/storage tech.					
1 (not knowledgeable at all)	368	0.18	0.38	0	1
2	368	0.38	0.49	0	1
3	368	0.36	0.48	0	1
4	368	0.07	0.25	0	1
5 (extremely knowledgeable)	368	0.01	0.09	0	1
Age Group					
18-30	368	0.54	0.5	0	1
31-40	368	0.29	0.46	0	1
41-50	368	0.15	0.36	0	1
51-60	368	0.02	0.13	0	1

Table 5. Correlation Table.

Variables	Mean	Std. Dev.	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
(1) Perceived Similarity	3.48	1.22	1															
(2) Topic Modeling Similarity	0.17	0.37	0.007	1														
(3) Blended algorithm Similarity	0.17	0.37	0.408	-0.201	1													
Gender																		
(4) Male	0.54	0.5	-0.073	-0.017	0.019	1												
(5) Female	0.42	0.49	0.055	0.009	-0.011	-0.92	1											
(6) Non-binary	0.02	0.15	0.076	-0.023	0.023	-0.173	-0.134	1										
(7) Did not disclose	0.02	0.13	-0.02	0.058	-0.058	-0.14	-0.109	-0.02	1									
Knowledge about battery/storage tech.																		
(8) 1 (not knowledgeable at all)	0.18	0.38	-0.049	0.02	-0.021	-0.211	0.252	-0.074	-0.06	1								
(9) 2	0.38	0.49	0.011	-0.006	0.004	-0.153	0.118	0.092	0.031	-0.368	1							
(10) 3	0.36	0.48	0.019	0.012	-0.009	0.251	-0.26	-0.01	0.036	-0.354	-0.596	1						
(11) 4	0.07	0.25	-0.005	-0.029	0.028	0.11	-0.089	-0.042	-0.034	-0.123	-0.208	-0.2	1					
(12) 5 (extremely knowledgeable)	0.01	0.09	0.06	-0.04	0.04	0.083	-0.076	-0.014	-0.012	-0.042	-0.071	-0.069	-0.024	1				
Age Group																		
(13) 18-30	0.54	0.5	0.07	0.003	-0.005	-0.193	0.118	0.147	0.119	-0.078	-0.01	0.101	-0.02	-0.098	1			
(14) 31-40	0.29	0.46	0.023	-0.014	0.013	0.123	-0.071	-0.102	-0.083	-0.021	0.044	-0.078	0.047	0.141	-0.696	1		
(15) 41-50	0.15	0.36	-0.144	0.015	-0.009	0.16	-0.127	-0.067	-0.055	0.157	-0.1	-0.006	-0.02	-0.038	-0.457	-0.273	1	
(16) 51-60	0.02	0.13	0.05	0	-0.001	-0.14	0.153	-0.02	-0.017	-0.06	0.163	-0.097	-0.034	-0.012	-0.139	-0.083	-0.055	1

No. of observations 368. The correlations above that absolute value of .12 are significant at $p<.05$ level.

With these answers, we evaluate the following regression model:

$$Perceived\ similarity = \beta 1 * Topic\ modeling\ similarity + \beta 2 * Blended\ algorithm\ similarity + Controls + constant$$

The dependent variable is the *perceived similarity* measured on a 5-point Likert scale (1 being the lowest and 5 being the highest), capturing the similarity assessments of the subjects for the patent pairs. The independent variables - *Topic modeling similarity and Blended algorithm similarity* are the two manipulations on which we selected the patent pairs. *Topic modeling similarity* is a binary variable with 1 representing high similarity and 0 as low similarity assessments made by Topic modeling between the patent pairs. Similarly, *Blended algorithm similarity* is also a binary variable, with 1 representing high similarity and 0 as low similarity assessments made for patent pairs by our *Blended approach* .

We control for the *Gender* (captured on four levels - Male, Female, Non-Binary, and Not disclosed), *Knowledge about batter technology* (captured on 5 points Likert scale with five as the highest level of knowledge), and *Age* (captured on four levels between 18 and 60 years) of the participants. The controls are all self-reported variables. We run an OLS regression to evaluate the coefficients relating topic modeling and *Blended algorithm* similarity assessments with perceived similarities. We also split regressions to assess the effects of two manipulations (as in experiment 1 and experiment 2) separately. And we find similar results[7].

Table 6 presents the regression results. Since perceived similarity is an ordinal variable, we also run a Generalized ordered logit model, and the results remain the same.

[7] The regression results are logged in Appendix that is available on request.

Table 6. Regression results.

	Perceived Similarity
VARIABLES	(1)
Topic Modeling Similarity	0.318*
	(0.171)
Blended algorithm Similarity	1.382***
	(0.175)
Gender (base Did not disclose)	
Male	0.006
	(0.558)
Female	0.166
	(0.560)
Non-binary	0.558
	(0.588)
Knowledge about battery/storage tech. (base 1-Not knowledgeable at all)	
2	0.048
	(0.151)
3	0.149
	(0.160)
4	0.054
	(0.281)
5 (extremely knowledgeable)	0.750***
	(0.200)
Age Group (base 18-30)	
31-40	-0.012
	(0.143)
41-50	-0.419**
	(0.172)
51-60	0.374
	(0.278)
Constant	2.129***
	(0.568)
Observations	368
R-squared	0.207

We observe a partially significant positive coefficient of the *Topic modeling similarity* ($\beta 1$ =0.318, p = 0.063) and a significantly positive coefficient of the *Blended algorithm similarity* ($\beta 2$ = 1.382, p = 0. 000). The regression coefficient $\beta 2$ has a higher magnitude than $\beta 1$. Notably, post hoc power analysis conducted with G*POWER (Faul, Erdfelder, Lang, & Buchner, 2007) revealed acceptable power measures of 0.97 or higher, suggesting that the size of our participant group is sufficient for the analyses (Cohen, 1992). We also need to test if the coefficients are significantly different. To test the equality of two coefficients (H0: $\beta 1 = \beta 2$), we perform the Wald test[8]. The Wald test (H0: $\beta 1 = \beta 2$) result is significant, with F=21.99, $\Delta df = 1$, $p < .05$ (p value=0.000), suggesting that the similarities assessments by human subjects relate to topic modeling and *Blended algorithm* similarity assessments to different degrees. We find that human subjects perceive similarity significantly higher alignment with our *Blended algorithm* than Topic modeling similarity assessments.

CONCLUSION AND FUTURE DIRECTIONS

Patents are considered valuable firm resources and are strategically used by firms to gain a competitive advantage. Patent statistics as proxy measures of a firm's innovation activities and performance have been central to empirical research. However, assessing the value of a patent is difficult as it depends on various factors. Moreover, patents do not necessarily require disclosure of technology performance targets and assessments, which could bias our understanding of the economic valuation of patents and targets in technology-led acquisitions. To address this issue, we leverage the differences between R&D activity entailing patents and academic publications to reveal performance targets of patented technology. We devise a blended approach that maps the knowledge disclosed in academic publications to patented inventions and helps reveal performance

[8] Refer to https://www.stata.com/manuals/rtest.pdf for more on Wald test.

targets. Our devised technique can be instrumental in understanding hidden paths in corporate research. We test our technique in the context of electric energy storage research for electric vehicles, and our blended approach performs significantly better than the topic modeling on patents alone. Our approach can be improved upon, given the fast-paced AI developments keeping the base idea the same. It can also be applied to other technological areas to reveal performance targets of patented technology.

Machine learning applied to large data sets holds significant promise in various problems. The blended approach presented in this study can disclose hidden competition among firms and interpret better competitive dynamics and actions. Our contribution is one of the first machine-learning-based methods to address the omission of performance targets in patent documents. The proposed algorithm can yield results that significantly impact information transparency in the patent process and market valuation of patents. More importantly, such results can help address the issue of long-term risks in merger and acquisition deals. Obtaining a more profound view of a firm's "undisclosed" innovation could reduce uncertainty and allow decision-makers to make more informed decisions. Finally, using electric mobility as our application domain, we can also offer insights to energy policymakers regarding projections about electric mobility adoption. A full electric mobility adoption is expected to have multiple societal, environmental as well as financial benefits. Hence, it is placed high in policy agendas. Thus, our algorithm can offer a more forward-looking view to policymakers toward designing future electric mobility policies. We hope this study fosters new directions for future research. For example, it would be interesting to replicate this methodology in technological landscapes characterized by different levels of uncertainty or exploratory trial-and-error approaches.

REFERENCES

Alcacer, J., & Gittelman, M. 2006. Patent citations as a measure of knowledge flows: The influence of examiner citations. ***The Review of Economics and Statistics***, 88(4): 774–779.

Altman, N. S. 1992. An introduction to kernel and nearest-neighbor nonparametric regression. ***The American Statistician***, 46(3): 175–185.

Aras, H., Hackl-Sommer, R., Schwantner, M., & Sofean, M. 2014. Applications and Challenges of Text Mining with Patents. ***IPaMin@ KONVENS***.

Arora, A., & Gambardella, A. 1994. The changing technology of technological change: General and abstract knowledge and the division of innovative labour. ***Research Policy***, 23(5): 523–532.

Barney, J. 1991. Firm resources and sustained competitive advantage. ***Journal of Management***, 17(1): 99–120.

Blei, D. M. 2012. Probabilistic topic models. ***Communications of the ACM***, 55(4): 77–84.

Blei, D. M., Ng, A. Y., & Jordan, M. I. 2003. Latent dirichlet allocation. ***Journal of Machine Learning Research***, 3(Jan): 993–1022.

Campbell, E. G., Weissman, J. S., Causino, N., & Blumenthal, D. 2000. Data withholding in academic medicine: Characteristics of faculty denied access to research results and biomaterials. ***Research Policy***, 29(2): 303–312.

Cano, Z. P., Banham, D., Ye, S., Hintennach, A., Lu, J., et al. 2018. Batteries and fuel cells for emerging electric vehicle markets. ***Nature Energy***, 3(4): 279–289.

Cao, J., Xia, T., Li, J., Zhang, Y., & Tang, S. 2009. A density-based method for adaptive LDA model selection. ***Neurocomputing***, 72(7–9): 1775–1781.

Cohen, J. 1992. Statistical power analysis. ***Current Directions in Psychological Science***, 1(3): 98–101.

Cover, T., & Hart, P. 1967. Nearest neighbor pattern classification. ***IEEE Transactions on Information Theory***, 13(1): 21–27.

Deng, Z., Zhu, X., Cheng, D., Zong, M., & Zhang, S. 2016. Efficient kNN classification algorithm for big data. ***Neurocomputing***, 195: 143–148.

Dinger, A., Martin, R., Mosquet, X., Rabl, M., Rizoulis, D., et al. 2010. Batteries for electric cars: Challenges, opportunities, and the outlook to 2020. ***The Boston Consulting Group***, 7: 2017.

Dosi, G. 1982. Technological paradigms and technological trajectories: A suggested interpretation of the determinants and directions of technical change. ***Research Policy***, 11(3): 147–162.

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. 2007. G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. ***Behavior Research Methods***, 39(2): 175–191.

Gambardella, A., Harhoff, D., & Verspagen, B. 2008. The value of European patents. ***European Management Review***, 5(2): 69–84.

Gans, J. S., Hsu, D. H., & Stern, S. 2008. The impact of uncertain intellectual property rights on the market for ideas: Evidence from patent grant delays. ***Management Science***, 54(5): 982–997.

Gittelman, M. 2008. A Note on the Value of Patents as Indicators of Innovation: Implications for Management Research. ***Academy of Management Perspectives***, 22(3): 21–27.

Golembiewski, B., Vom Stein, N., Sick, N., & Wiemhöfer, H.-D. 2015. Identifying trends in battery technologies with regard to electric mobility: Evidence from patenting activities along and across the battery value chain. ***Journal of Cleaner Production***, 87: 800–810.

Griffiths, T. L., & Steyvers, M. 2004. Finding scientific topics. ***Proceedings of the National Academy of Sciences***, 101(suppl_1): 5228–5235.

Guo, G., Wang, H., Bell, D., Bi, Y., & Greer, K. 2006. Using k NN model for automatic text categorization. ***Soft Computing***, 10: 423–430.

Hall, B. H., Jaffe, A., & Trajtenberg, M. 2005. Market Value and Patent Citations. ***The RAND Journal of Economics***, 36(1): 16–38.

Hessels, L. K., & Van Lente, H. 2008. Re-thinking new knowledge production: A literature review and a research agenda. ***Research Policy***, 37(4): 740–760.

Jaffe, A. B., & Trajtenberg, M. 2002. ***Patents, citations, and innovations: A window on the knowledge economy***. MIT press.

Jeppesen, L. B., & Lakhani, K. R. 2010. Marginality and problem-solving effectiveness in broadcast search. ***Organization Science***, 21(5): 1016–1033.

Lanjouw, J. O., & Schankerman, M. 2001. Characteristics of patent litigation: A window on competition. ***RAND Journal of Economics***, 129–151.

Lim, M. K., Mak, H.-Y., & Rong, Y. 2015. Toward mass adoption of electric vehicles: Impact of the range and resale anxieties. ***Manufacturing & Service Operations Management***, 17(1): 101–119.

Ma, S.-C., Xu, J.-H., & Fan, Y. 2022. Characteristics and key trends of global electric vehicle technology development: A multi-method patent analysis. ***Journal of Cleaner Production***, 338: 130502.

Masias, A., Marcicki, J., & Paxton, W. A. 2021. Opportunities and challenges of lithium ion batteries in automotive applications. ***ACS Energy Letters***, 6(2): 621–630.

Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. 2013. Distributed representations of words and phrases and their compositionality. ***Advances in Neural Information Processing Systems***, 26.

Nordhaus, W. D. 1972. The optimum life of a patent: Reply. ***The American Economic Review***, 62(3): 428–431.

Perneger, T. V., & Hudelson, P. M. 2004. Writing a research article: Advice to beginners. ***International Journal for Quality in Health Care***, 16(3): 191–192.

Peteraf, M. A. 1993. The cornerstones of competitive advantage: A resource-based view. ***Strategic Management Journal***, 14(3): 179–191.

Pollet, B. G., Staffell, I., & Shang, J. L. 2012. Current status of hybrid, battery and fuel cell electric vehicles: From electrochemistry to market prospects. ***Electrochimica Acta***, 84: 235–249.

Rauh, N., Franke, T., & Krems, J. F. 2015. Understanding the impact of electric vehicle driving experience on range anxiety. ***Human Factors***, 57(1): 177–187.

Reitzig, M. 2003. What determines patent value? Insights from the semiconductor industry. ***Research Policy***.

Rosenberg, N. 2010. Why do firms do basic research (with their own money)? ***Studies on science and the innovation process: Selected works of Nathan Rosenberg***: 225–234. World Scientific.

Rotolo, D., Camerani, R., Grassano, N., & Martin, B. R. 2022. Why do firms publish? A systematic literature review and a conceptual framework. ***Research Policy***, 51(10): 104606.

Salandra, R. 2018. Knowledge dissemination in clinical trials: Exploring influences of institutional support and type of innovation on selective reporting. ***Research Policy***, 47(7): 1215–1228.

Sampat, B., & Williams, H. L. 2019. How do patents affect follow-on innovation? Evidence from the human genome. ***American Economic Review***, 109(1): 203–236.

Sbalchiero, S., & Eder, M. 2020. Topic modeling, long texts and the best number of topics. Some Problems and solutions. ***Quality & Quantity***, 54: 1095–1108.

Schankerman, M., & Pakes, A. 1986. Estimates of the Value of Patent Rights in European Countries During the Post-1950 Period. ***The Economic Journal***, 96(384): 1052–1076.

Somaya, D. 2012. Patent Strategy and Management: An Integrative Review and Research Agenda. ***Journal of Management***, 38(4): 1084–1114.

Trappey, A. J., Trappey, C. V., Wu, C.-Y., & Lin, C.-W. 2012. A patent quality analysis for innovative technology and product development. ***Advanced Engineering Informatics***, 26(1): 26–34.

Valogianni, K., Ketter, W., Collins, J., & Zhdanov, D. 2020. Sustainable electric vehicle charging using adaptive pricing. ***Production and Operations Management***, 29(6): 1550–1572.

Chapter 2: Regulatory disclosures and competitive signals: Competition affecting speed and tenor of clinical trial reporting

ABSTRACT

Regulatory disclosure mandates require firms to share information and can exacerbate competitive pressures. However, the extant literature views regulatory compliance as a burden for firms that abide reluctantly, overlooking competitive dynamics and signaling. Combining insights from the literature on regulatory disclosures and competitive signaling theory, we propose that when regulatory disclosure mandates pit firms against each other, they respond to the magnified competitive pressures by swift information sharing and adopting a more positive tenor in their disclosures. Furthermore, firms with a reputation for introducing novel products display more urgency and fervor to signal to competitors and reduce uncertainty about their ongoing product developments. We test our hypotheses in the pharmaceutical setting, leveraging a multi-source dataset comprising clinical trials completed between 2008 and 2019 and registered on ClinicalTrials.gov under the ambit of FDAAA 801 regulation. We find overall support for our predictions and discuss the implications of our findings for theory and practice.

INTRODUCTION

Competitive pressures can induce firms to use information signals (Connelly, Certo, Ireland, & Reutzel, 2011; Heil & Robertson, 1991; Livengood & Reger, 2010) and to communicate strategically (Gao, Yu, & Cannella, 2016; Guo, Sengul, & Yu, 2020; Guo, Yu, & Gimeno, 2017). Information signals—"conduct and observable attributes that alter the beliefs of, or convey information to, other individuals in the market about unobservable attributes and intentions" (Ndofor & Levitas, 2004, p. 688)—can alter competitors' beliefs about firm intentions (e.g., finding cures for a disease condition, in our context) and discourage new entrants (Clarkson & Toh, 2010). As firms navigate a competitive landscape, they may respond through vague communication (Guo et al., 2017), aggressive "Keep Out signs" (Clarkson & Toh, 2010), and announcing new products, some of which may be a bluff (e.g., vaporware: Bayus, Jain, & Rao, 2001; Dranove & Gandal, 2003).

However, most of the literature on the strategic use of information considers voluntary disclosure (Lev, 1992; Levitas, McFadyen, & Ahsan, 2016; Marquis, Toffel, & Zhou, 2016; Stern & James, 2016; Toh & Miller, 2017), largely overlooking strategic maneuvering in regulatory contexts. Further, the dominant view on regulatory disclosure considers these mandates a burden for firms, such that firms comply reluctantly (Desai, 2016; Fabrizio & Kim, 2019; Oliver, 1991; Shi & Connelly, 2018). Firms use their influence to thwart regulatory mandates (Ahuja & Yayavaram, 2011; Zhou, 2022), attempt symbolic compliance over substantive action (Edelman, Uggen, & Erlanger, 1999; Oliver, 1991), or delay compliance to reduce institutional complexity (Raaijmakers, Vermeulen, Meeus, & Zietsma, 2015). This resistance is justified for firms facing competition as they fear knowledge leakages and have appropriability concerns (Cohen, Goto, Nagata, Nelson, & Walsh, 2002; Devarakonda & Reuer, 2018; James, Leiblein, & Lu, 2013; Nelson, 2016; Winter, 2006).

Regulatory mandates, or "rules of the game" (North, 1990, p. 3), can exacerbate competitive pressure and affect a firm's incentives to disclose proprietary information. First, regulatory mandates increase managerial awareness of the competitive environment by making reliable and contextual information available about the activities of other firms. This additional awareness about other firms' activities makes the competition salient. The information shared on the regulatory platform places firms with overlapping projects in proximity to each other and tangibly in a race to finish (Chen, Su, & Tsai, 2007; Porac, Thomas, & Baden-Fuller, 1989). At the same time, the regulatory forum renders sanctity to firm actions, such that firms will actively avoid any unlawful conduct. Regulatory oversight and scrutiny make firm actions more credible to an observer. For instance, a firm cannot bluff its competitors or underplay negative information if a regulator requires "hard" information in set templates (Liberti & Petersen, 2018; Tian, 2015). This interplay between regulatory disclosure demands and firms' competitive environment is particularly salient in knowledge-based industries such as pharmaceuticals and biotech, which are heavily regulated and in which information about the R&D activities of firms can shift the competitive landscape (Caner, Bruyaka, & Prescott, 2018; Grant, 1996; James et al., 2013; Mc Namara & Baden-Fuller, 2007; Ndofor & Levitas, 2004).

In this study, we examine the assumption that firms comply reluctantly with regulatory disclosures and ask whether disclosure mandates prompt firms to respond strategically and signal. We bring insights from the competitive signaling theory (Connelly et al., 2011; Gao et al., 2016; Guo et al., 2017; Heil & Robertson, 1991) to regulatory disclosures (Aghamolla & Thakor, 2022;

Aragon-Correa, Marcus, & Vogel, 2020; Leuz & Wysocki, 2016; Mahoney, 1995; North, 1990) to model a firm's disclosure behavior on a regulatory platform.

We argue that the perception of high competition due to regulatory disclosures exacerbates the competitive threat to firms, making them resort to competitive signaling and being prompt and positive in their disclosures. While regulatory guidelines restrain mandated content, firms use the disclosure speed and tenor of voluntarily shared information as a signal. The swift sharing of the required information and positive tenor of associated voluntary content convey competence and commitment to competitors. Further, a firm's prior success in navigating regulatory processes and introducing novel products builds a reputation (Dimov, Shepherd, & Sutcliffe, 2007; Love & Kraatz, 2009; Pfarrer, Pollock, & Rindova, 2010), signaling the firm's edge over competitors. As the signaling cost is small for a reputed firm ("separating equilibrium" posits that high-quality senders incur low costs for credible signals: Bergh, Connelly, Ketchen, & Shannon, 2014; Connelly et al., 2011), its managers are prompt and positive in disclosing information; else, external parties, including competitors, might perceive delays or unflattering disclosures as a signal of bad news (Grossman, 1981; Milgrom, 1981; Sheth, 2021). Thus, firms with a high reputation will seek signal congruence and be more aggressive in their response to the competitive threat (Bapna, 2019; Drover, Wood, & Corbett, 2018; Stern, Dukerich, & Zajac, 2014).

We explore these issues in the context of clinical trials undertaken by pharmaceutical firms that fall under the Food and Drug Administration (FDA) Amendments Act (FDAAA) 801, introduced in the United States (US) in 2007.[1] The FDAAA requires responsible parties to (i) register clinical trials publicly and (ii) share trial outcome information on ClinicalTrials.gov after completion. The registry enlists privately and publicly funded clinical studies conducted worldwide and is increasingly used to identify trends in clinical research and track competitors.

As a leading industry strategist opined, “ClinicalTrials.gov is extremely useful to monitor competitor clinical trial patient populations, timelines, endpoints, recruitment status, sites and general progression. The RSS feature is helpful in being alerted to any changes in trials of interest or if new trials are posted in a competing space (Booth, 2020).” While registering a study by a firm marks the initiation of a line of inquiry, publishing trial results carries a substantial cost and strategic value as a signal of headway, e.g., results suggest whether a drug is effective and safe for commercialization.[2] Furthermore, regulatory oversight marks this information with a low probability of bluffing (Bergh et al., 2014; Connelly et al., 2011; Gao et al., 2016; Heil & Robertson, 1991). Thus, the regulator-demanded disclosures of trial results provide a unique setting to observe firms’ disclosure behavior and test the prevalence of competitive signaling in a regulatory forum.

To test our theory, we put together a multi-source dataset that leverages information on (i) trials completed between 2008 and 2019 from ClinicalTrials.gov; (ii) trial sponsors from Compustat; (iii) approved drugs from the FDA Orange Book; and (iv) publications linked to the trial from PubMed. We also use MeSH terms to map the studies to the unique disease concepts, pre-trained clinical concept embeddings to map therapeutic areas to the trials, and text analysis to assess the publication tenor (i.e., whether the publications have a positive or negative valence).

This study aims to make several contributions to theory and practice. First, we test the assumption of a competition-induced firm’s reluctance to comply with regulatory demands for information disclosure. Our results suggest the presence of competitive signaling in a regulatory context. We drive insights from the competitive signaling theory to show that the regulatory forum emulates a competitive contest and induces firms to use disclosures (their speed and tenor) for signaling (Clarkson & Toh, 2010; Gao et al., 2016; Guo et al., 2017; Ndofor & Levitas, 2004).

Second, we offer evidence of firms' linguistic maneuvering in scientific publications responding to competitive pressures. In doing so, we contribute to understanding competitive contexts where knowledge is embedded in language, adding to competitive signaling (Guo et al., 2017; Heil & Robertson, 1991) and rhetoric signaling theories (Gutiérrez, Papiashvili, Tribó, & Vazquez, 2020; Steigenberger & Wilhelm, 2018).

Third, while reporting bias related to clinical trials has been examined in past work (DeVito & Goldacre, 2019; Dwan, Gamble, Williamson, & Kirkham, 2013; Salandra, 2018), we add to these studies to investigate the use of reporting speed and language as a strategic tool evident even in scientific reports (Gao et al., 2016). In doing so, we contribute to the debate around the efficiency and shortfalls of public clinical trial registries (Castellani, 2013; Dahm, Gonzalez, & Porteiro, 2009; Goldacre, 2013; Hsu, Lee, Moon, & Oh, 2019; Zarin & Tse, 2008). While policymakers might introduce regulations demanding information to repair trust among different stakeholders and enhance transparency, it makes firms more aware of overlapping competitive activities and drives a firm's propensity to signal. We delve into the broader implications of our study in the discussion section.

THEORY AND HYPOTHESES

Regulatory Disclosures and Competition

Economic models predict that, in the absence of disclosure costs, firms would be prompt to disclose complete information, lest the market would perceive any disclosure failure as a signal of bad news (Grossman, 1981; Milgrom, 1981; Sheth, 2021). However, if competitors could use disclosed information at the disclosing firm's expense, the potential proprietary costs of disclosures increase, decreasing incentives for disclosure. The proprietary costs of disclosures are particularly high in the context of R&D information mandates for firms facing high competition as they fear knowledge leakages and have appropriability concerns (Cohen et al., 2002;

Devarakonda & Reuer, 2018; Nelson, 2016; Winter, 2006). As a result, the literature suggests a reluctance to disclose any proprietary information, to the point where firms lobby against reporting mandates to protect their competitive position and avoid competitive harm (Ahuja & Yayavaram, 2011; Zhou, 2022). Firms also attempt symbolic compliance over substantive action (Edelman et al., 1999; Oliver, 1991), and delay compliance to reduce institutional complexity (Raaijmakers et al., 2015).

Regulatory mandates change the "rules of the game" and affect firms' strategic environments (Ahuja & Yayavaram, 2011; North, 1990). The information shared by firms affects how firms are seen by their stakeholders, potential buyers, and investors (Cade, Kaplan, & Loftus, 2021; Healy & Palepu, 2001; Jayaraman & Wu, 2019; Kajüter, Klassmann, & Nienhaus, 2019). They also shape managerial decision-making by impacting strategic decisions such as those around investments (Roychowdhury, Shroff, & Verdi, 2019) and going public. For example, Aghamolla and Thakor (2022) show that clinical trial disclosure requirements under FDAAA make going public relatively attractive for private firms because managers can no longer avoid proprietary disclosure costs by remaining private.

While the concern of competition driving a firm's resistance to regulatory disclosures is extensively documented in the literature, studies exploring the interaction of regulatory disclosure demands and competitive dynamics are scant, if not wholly absent. Regulatory demands shape the competitive dynamics between firms, as regulatory disclosures affect the information environment of the firm, increasing the firm's monitoring of competitors' activities, which is a fundamental concern in competition studies. The increased awareness about the presence of other firms in a market or product space enhances the competitive threat (Chen & Miller, 2012; Downing, Kang, & Markman, 2019). For example, as decision-makers make sense of their competitive environment by constructing cognitive models considering the available information (De Chernatony, Daniels, & Johnson, 1993; Porac et al., 1989), other firms' projects in the overlapping market account for the enhanced competitive perception. Firms interact with one another across stages of innovation in various spaces, such as publications and patents (Baker & Mezzetti, 2005; Lanjouw & Schankerman, 2001), product development and regulatory approvals (McCann & Bahl, 2017), and product marketing, with each phase and space having its own competitive presence. A unique feature of the regulatory forum (e.g., ClinicalTrials.gov) is that information is reported in the same repository and time frame, enhancing the spatial and temporal proximity and making competitors' presence salient. As a result, regulatory information requirements may impact a firm's competitive environment and disclosure behavior.

Competition, Competitive Signaling, and Regulatory Disclosures

Competition is, in turn, a strong motivator for firms to initiate and respond with strategic moves (Chen, 1996; Chen & Miller, 2012). In their seminal work, Heil and Robertson (1991) put forward the idea that firms signal proposed market actions to competitors by extending signaling theory in the context of competition. Firms send competitive market signals (i.e., announcements or previews of potential actions intended) to convey or seek information from competitors. The signals can be substantive or rhetorical (Steigenberger & Wilhelm, 2018), truthful or bluffs—carrying varying levels of credibility (Haan, 2003), and are anchored in the signaling firm's intentions (Heil & Robertson, 1991). Firms can also bundle complementary signals to convey information about different aspects. The risk/benefit tradeoff dictates a firm's signaling propensity, such that signals will either discourage potential competitors from following or the signaling firm fears that revealing intentions to competitors might alert competitors who can initiate spoiler actions in the marketplace (Heil & Robertson, 1991; Lev, 1992). Therefore, the likelihood of signaling is contingent upon managerial perceptions of the firm's competitive environment (Ndofor & Levitas, 2004). The propensity to signal is also context-dependent, considering potential spillovers and appropriability conditions (Jansen, 2010).

Competitive signaling theory suggests that the intention behind a signal—a potential market action—should be deliberate (Heil & Robertson, 1991). Firm discretion on information content is loosened due to regulatory disclosure mandates. For instance, the FDAAA requires firms to submit summary results of their "applicable clinical trials" (ACTs) in a specified template no later than 12 months after the primary completion date (i.e., one year after the final participant is examined or receives an intervention). The results include administrative details and scientific information in a tabular format without the narrative or rhetoric about the study outcomes that might be seen in a scientific publication (Tse, Williams, & Zarin, 2009). However, study investigators can voluntarily attach scientific publications supporting protocol choices and trial findings, leaving some firm discretion.

The disclosure mandates also alter a firm's information environment by increasing the flow of information about competing firms' activities. As the information available increases, so does the firm's apprehension by making competition salient. The regulatory forum also lends credibility to the information reported on it. For instance, in the context of ClinicalTrials.gov, the regulatory oversight and quality checks ensure the sanctity of the posted information. The rule-based system assures the information quality during the data entry stage, and then the NIH staff manually reviews the quality. The reported results include core study outcomes, which allow for assessing trial success or failure, as well as information about the statistical significance of the findings (Tse et al., 2009). Therefore, the regulatory mandates increase the incentives and reduce the additional costs of disclosures, rendering the signaling mechanism more attractive.

Competitive Signaling and Speed of Disclosure

So far, we have argued that the regulatory forum makes signaling mechanisms seem more attractive to firms as the mandates reduce the marginal cost of disclosure, accentuate competitive threats, and render certain reliability to information shared on the forum. In the competitive realm, speed is a strategic concern involving managerial deliberations on associated costs and benefits (Blackburne & Quinn, 2020) and serves as a signal of commitment (Heil & Robertson, 1991). Firms respond to competitive threats with speedy actions (Chen & Hambrick, 1995; Nadkarni, Pan, & Chen, 2019). We expect that perceived competition in the context of regulatory mandates will induce firms to disclose faster because speedier disclosures signal competency and commitment to external stakeholders. As regulatory templates constrain firms' control over shared information and limit their signaling repertoire, firms tend to use information sharing speed as a signal. In our context, prompt results sharing can also help differentiate the focal firm from competitors to the FDA, which gives the final nod for the product launch. This can culminate in lead time advantages for the focal firm and affect consumer preferences (Carpenter & Nakamoto, 1989; Heil & Robertson, 1991; Lieberman & Montgomery, 1988). The first-mover advantage is particularly important in the pharmaceutical industry, as the first-entrant still has a higher market share than the second-entrant even many years after launch (Cha & Yu, 2014; Grabowski & Vernon, 1992). The prompt availability of the results also encourages prospective subjects to enroll in subsequent firm trials, speeding the process of gathering subsequent clinical evidence (Caldwell, Hamilton, Tan, & Craig, 2010).

Second, we expect firms to signal product development headway by speedier disclosure to ward off competition. For instance, firms preannounce new products or sometimes announce products that are never released (e.g., vaporware) to deter competitors (Bayus et al., 2001; Dranove

& Gandal, 2003). Clarkson and Toh (2010) show this deterrence effect, or "Keep Out sign," through patent reexamination certificates issued to firms in the US communication equipment industry. In the context of ClinicalTrials.gov, the disclosure relates to trials submitted to regulators, making them more credible signals than firms that employ a vaporware strategy. Taken together, we expect firms under regulatory mandates and intense competition to hurry with disclosures of results. Therefore,

> *Hypothesis 1: Competition positively affects the speed of disclosures when firms disclose information in a regulatory forum.*

Competitive Signaling and Disclosure Tenor

Until now, we have argued that competition induces firms to increase the speed of disclosure in a regulatory forum. We now turn to consider the content of the disclosures, and specifically, their tenor. We argue that firms will use their information advantage to describe their studies in the most positive light. Wherever possible, firms will relate positive tenor content as a signaling response to high competition threats. Competitive signaling theory has increasingly engaged with the 'language games' (Gao et al., 2016; Guo et al., 2017) and, more recently, revisited its "cheap talk" assumption (Stein, 1989) that sending a signal is not costly for the sender. For example, Steigenberger and Wilhelm (2018) extend the signaling theory to show that firms employ rhetorical signals to gain stakeholder support, and Guo and colleagues (2017) show that firms use language to deter new entrants.

While language games are commonly lamented in corporate communications issued by firms (e.g., Graffin, Haleblian, & Kiley, 2016; Guo et al., 2020, 2017; Westphal & Bednar, 2008), academic publications are traditionally seen as scientific endeavors, free from any language manipulations. However, evidence increasingly supports linguistic imprints in scientific publications (Pennebaker, 2011; Tausczik & Pennebaker, 2010). For instance, Markowitz and

Hancock (2016) found that scientific publications retracted for fraudulent data used more linguistic obfuscation than unretracted papers. On ClinicalTrials.gov, the regulatory template restricts firms from using rhetoric or "soft information" in their disclosures.[3] As these restrictions constrain the firms' control over shared information and limit their signaling repertoire; firms tend to attach voluntary information, such as references to scientific publications.

Scientific publications are peer-reviewed, costly to produce, and 'scientific' in nature, and, thus, a more credible signal than other dissemination channels (e.g., press releases). These publications are an important vehicle for market dominance. For example, in clinical research, they influence physicians' prescription choices over rival treatments (Azoulay, 2002). In the same way that competition induces firms to publish scientific articles and position innovations relative to competitors (Polidoro & Theeke, 2012), we expect competition to create inducements to use a more positive tenor.

Furthermore, publications guide the interpretation of the results, conveying a firm's take on the findings and its commitment to the further product development. This is particularly important in the context of complex studies. In clinical trials, for example, information is difficult to interpret even by experts, leaving the scope for interpretation (misinterpretation) bias. Wu and colleagues (2016) found that ClinicalTrials.gov trial descriptions are even more challenging to read than clinician notes and require years of formal education to understand. Therefore, firms can inculcate confidence in their project choices and outcomes by referencing scientific publications with a positive tenor and giving a positive spin to the reported results data. To summarize, we propose that high perceived competition will induce firms to support their mandatory disclosures by attaching publications with a positive tenor. Therefore,

Hypothesis 2: Competition positively affects the tenor of disclosures when firms disclose information in a regulatory forum.

Reputation and Competitive Signaling

Here, we further build on the competitive signaling mechanism in the regulatory disclosure forum and consider the effect of a firm's reputation on the theorized relationships between competition and disclosure elements. Reputation is a form of "technical efficacy" built by repeated action and success in the firm's efforts and is "tightly coupled to consequences and tangible organizational outputs (Love & Kraatz, 2009, p. 317)." It helps reduce uncertainty by signaling future performance based on observable past performance (Dimov et al., 2007).

In the pharmaceutical industry, NDAs (New Drug Approvals)—"an active ingredient that has never before been marketed in the United States in any form" (FDA, 2021)—owned by the firm, render a reputation for innovation and regulatory competence to the firm. Securing NDAs entails a long (on average, ten years) and uncertain process (with a single-digit success rate) for any firm (Scherer, 2010). It involves managing multiple trials and submitting voluminous test evidence to support its application. Therefore, successful NDAs convey a form of competence in a firm's clinical R&D efforts. A firm's prior success in FDA drug approvals also suggests the firm's superior ability to navigate the regulatory process.

Managers infer a firm's ability as a credible defender of its market interests through its previous success (Clark & Montgomery, 1998). While a high reputation signals a firm's capabilities in successfully maneuvering the R&D process, it does not convey much about the quality of the individual project, such as a drug being tested in a trial. Thus, we expect high-reputation firms to reduce uncertainty about their project outcomes for their competitors through speedy disclosures and positive rhetoric. In other words, firms will complement their signals with their earned reputation to convey their superior attributes to their competitors (Bapna, 2019; Dimov et al., 2007). A high-reputation signal coupled with prompt disclosures and positive rhetoric will offer a double shield to firms against competitors due to signal congruence (Stern et al., 2014). When signals are congruent (i.e., reputation and trial disclosures are prompt and with a positive tenor), signal receivers confirm their judgments and expectations about the sender's abilities. When signals are incongruent, receivers process information piecemeal, with their expectations and the sender's attributes diverging (Drover et al., 2018). As the costs of disclosures are relatively small for a high-reputation firm, its managers are prompt and positive in disclosing information; or else, external parties, including competitors, might perceive disclosure delays as a signal of bad news (Grossman, 1981; Milgrom, 1981; Sheth, 2021). Thus, high-reputation firms will seek the benefit of signal congruence and avoid incongruency by being prompt and positive in their disclosures (Bapna, 2019; Drover et al., 2018; Stern et al., 2014). Therefore,

> *Hypothesis 3a: The firm's reputation positively moderates the relationship between competition and the speed of disclosures when firms disclose information in a regulatory forum.*

> *Hypothesis 3b: The firm's reputation positively moderates the relationship between competition and the tenor of disclosures when firms disclose information in a regulatory forum.*

METHODS

Sample and Data Sources

We drew our sample from the ClinicalTrials.gov registry. Well-documented biases in the publication and reporting of clinical trial outcomes have resulted in a general lack of trust in medical evidence claims and increased scrutiny of systematic errors in medical publications (Dwan et al., 2013). In the US, the government established ClinicalTrials.gov to provide greater transparency and improve the reliability of published science.[4] While the website was made available in 2000, reporting of clinical trials and results remained limited and largely voluntary until FDAAA 801 was enacted in 2007. Firms worry that their competitors would misrepresent their knowledge and have appropriability concerns, especially for early-stage trials (phase I), which are kept out of the law's ambit (DeVito, Bacon, & Goldacre, 2020; Goldacre, 2014). The FDAAA requires the "applicable clinical trials" (ACTs) to be registered within 21 days of enrolling the first participant and the submission of summary results no later than 12 months after the primary completion date (i.e., one year since the final participant is examined or receives an intervention). The results include administrative details and scientific information in a tabular format without the narrative or rhetoric about the study outcomes that might be seen in a scientific publication (Tse et al., 2009).

To understand what information is reported about a given trial, consider the trial with the identifier NCT01999322, which investigates a Type 4 diabetes drug.[5] The corresponding ClinicalTrials.gov record gives the tracking data such as the inception date, date of completion, date of results posted, and a history of the changes made to the trial record. The descriptive information section includes details on the disease condition, study design, intervention, and adverse events during the trials. The investigators can voluntarily provide academic studies linked to the trials. The section on recruitment has information on the study participants' sample size, eligibility criteria, and demographics. The final administrative information section has data about the study sponsor and collaborators.

We downloaded all registered trials on the platform until December 2019 (325,860 trials). The FDAAA 801 came into effect on September 27th, 2007, and applied to ongoing trials or those started after. As we were interested in the disclosure of trial results, we focused only on completed trials, leaving out ongoing or terminated (i.e., interrupted) trials. To identify trials that come under the purview of FDAAA, we followed the legislation's logic, including clinical trials of an FDA-regulated drug, biological, or device product other than phase I (for drug or biological products) or small feasibility studies (for device products) and a primary completion date of "January 2008" or later (Anderson et al., 2015; Tse et al., 2009).[6] This left us with 19,944 trials. Since our study focuses on the firms' strategic behavior, we identified studies that a firm ran. To do so, we leveraged the registry classification of a trial sponsor as 'Industry' if the lead sponsor is a firm, 'Non-industry' for hospitals and non-profit organizations, and 'US Government' for NIH and other federal agencies. This procedure left us with 11,053 trials. Next, we matched the trial sponsors in the Compustat database for public firms to attach firm-level information. The final sample includes 4,938 trials conducted by 396 public firms.

We then matched the firms in our sample to the Orange Book, a list of all FDA-approved products, to find their portfolio of marketed products. To map the disease conditions of the trials in therapeutic areas, we used the National Library of Medicine's (NLM) Unified Medical Language System (UMLS) database maintained by NIH. Finally, the sponsors can voluntarily include links to scholarly articles related to trials in the registry. We were able to link our dataset to 4,071 publications from 584 trials. To analyze the text of these publications to create the tenor variable, we downloaded them from the PubMed.gov database.[7]

Dependent Variables

Our first dependent variable is the *Disclosure Delay,* which we measured as the days intercurrent between the completion of the focal trial and the reporting of its results. The higher this measurement, the larger the delay in posting the results since trial completion.[8] Our second dependent variable is the *Disclosure Tenor.* For those trials where the results contained links to published studies, we scraped the text of academic papers and conducted a text analysis using LIWC 2015 software (Pennebaker, Booth, Boyd, & Francis, 2015). The software counted the total positive and negative affective words for each publication. Following previous literature (Quigley, Hubbard, Ward, & Graffin, 2020; Zavyalova, Pfarrer, Reger, & Hubbard, 2016), we measured the tenor of these publications as:

$$\textit{Publication Disclosure Tenor} = \frac{\textit{positive words}}{\textit{(positive words + negative words)}}$$

To calculate disclosure tenor at the trial level, we averaged the tenor measure across all studies linked to the trial. The higher this measurement, the more positive the tenor of the publications linked to a trial.

Independent Variables

Our main independent variable is the level of Competition faced by each trial sponsor. Firms engaged in drug development consider indication contact (i.e., developing a treatment for the same medical condition) as a critical factor in defining competition (Trapido, 2013). As spatial and temporal proximity affect competition perception and firm response (Litz & Pollack, 2015; Porac et al., 1989), other firms' projects in the overlapping medical condition and reported in the same regulatory forum (spatial proximity) in the same time frame (temporal proximity) will lead to enhanced competitive perception. Thus, we operationalized competition for the focal clinical trial CT_i as the number of unique clinical trials (i) which have at least one common disease/condition with CT_i and (ii) which are also overlapping in time with CT_i, i.e., they were started before the completion of CT_i and completed after the start of the CT_i. The National Library of Medicine (NLM) evaluates study data and compiles a list of standard Medical Subject Heading (MeSH) terms that describe the condition(s) addressed by each clinical trial. We used standard medical terms to identify projects with a common disease or condition.[9] For example, if CT_i had four condition terms matched with two other trials, such that overlapping Trial 1 had two matched conditions with CT_i and overlapping Trial 2 had all four matched conditions, the competition for CT_i was calculated as:

$$\text{Competition for } CT_i = \frac{2}{4} + \frac{4}{4} = 1.5$$

To test our moderating hypotheses, we proxied the firm's *Reputation* through their past record of New Drug Approvals (NDA), calculated as the number of NDAs that the firm possessed before the start of CT_i. As noted earlier, in our context, the prior success of the firm in FDA drug approvals suggests a reputation of superior attributes to intended receivers.

Control Variables

We controlled for several potential endogeneity sources that might influence the perceived competition and disclosing behavior. At the level of the trial, the number of sponsors and the resources they bring may be linked with both competition and reporting behavior. For example, well-resourced trials might be more common in competitive therapeutic areas. Thus, we controlled for the *Number of Sponsors* (including the lead sponsor) and the *Number of Facilities* used in a trial. We also controlled for trial size through the *Number of Subjects Enrolled* (this variable was log-transformed to account for its skew). Since later-stage trials might affect FDA approvals and have different signaling effects, we also controlled for the stage of the study by *Trial Phase* through a binary variable (1-Phase 3 or 4/0-Phase 1 or 2). As the reporting of a trial might be linked to whether its outcomes are favorable, we also controlled for the serious *Adverse Events* (percentage of adverse events out of total events) occurring during a trial. We consider this a proxy of the trial's success – the more serious events, the less favorable the trial outcomes. The trial sponsors can request an extension in posting their results to the FDA. As this might correlate with the delay in posting the results, we controlled for *Requested Extension* through a binary variable (1/0).

At the sponsor level, we controlled for *R&D Intensity* (R&D expenses scaled by the sales revenues before the year of focal trial start), *Firm Slack* (as captured by the ratio between current assets and current liabilities), and *Firm Size* (Log Assets) to account for knowledge capabilities.

To account for a firm's absorptive capacity that may affect the competition-signaling relationship, we also controlled for *Past Experience* (number of total previous trials undertaken by a sponsor).

Finally, as we wanted to explore the effects of competition over and above the effects of other types of competition, we controlled for product competition following insight from Ball, Shah, and Wowak (2018). We measure *Product Competition* as the ratio of Abbreviated New Drug Application (ANDA or applications for generic drugs) products actively marketed by a firm in a year divided by the total number of NDA and ANDA products marketed. The lower values suggest that the firm markets more NDA or specialized products and hence faces a lower threat of competition. Alternately, the higher ratio measure suggests that the firm markets more ANDA or generic products and faces intense competition across its portfolio of products. To calculate this measure, we collected and combined data from the FDA's Orange Book and lists of additions and deletions for prescription and OTC drug products.[10]

Since different therapeutic classes may attract different competitors and impact the result disclosures, we also controlled for the *Therapeutic Areas* under which the trials would fall. We followed the methodology of previous studies to map trials with therapeutic areas (Choi, Chiu, & Sontag, 2016; Si, Wang, Xu, & Roberts, 2019). Using the Unified Medical Language System (UMLS) database, we used MeSH terms to map each trial to the unique disease concepts. We then used pre-trained clinical concept embeddings made available by Vine and colleagues (2014) and derived from 348,566 medical-journal abstracts to map the associated therapeutic areas to the trials.[11] Appendix A summarizes the algorithm used to assign a therapeutic area to each study and the data sources used. Appendix B lists all the therapeutic areas along with the associated terms. Finally, we included *Year* dummy variables to control for time fixed-effects such as the increased awareness of the clinical trial registry and changes in regulations.

Estimation Methods

To understand how competition affects the speed and tenor of disclosure, we estimated both naïve and Heckman models (Heckman, 1979). Naïve models may be estimated for a non-random sample of trials, potentially biasing our results. Specifically, we can observe the speed of disclosure of trial results contingently upon the results being disclosed in the first place. Despite the legal obligation to report, approximately 67.5% of all the eligible trials in our set included reported results. The firms might have withheld the results for various reasons, so we cannot assume that the final sample is random. We thus use a two-step Heckman estimator (Certo, Busenbark, Woo, & Semadeni, 2016). As a robustness check, we report the naïve models in Appendix C.

In the first stage, we estimated the likelihood of disclosure. In the second step, we introduced the *Inverse Mills Ratio* to account for the possible selection bias. The Heckman model requires an exclusion restriction that is related to the selection (i.e., the likelihood of disclosure) but should not predict the outcome (i.e., speed and tenor of disclosure) except via its relationship in the first stage. We used the firm's *ratio of results reported to the total number of completed trials* in the same disease conditions as the identification variable. This variable reflects the firm's propensity for disclosure based on the firm's past disclosures and hence is highly correlated with the likelihood of disclosure of trial results but is not significantly associated with the speed or tenor of disclosure or the underlying error term of the second stage model. To evaluate whether our exclusion restriction was sufficiently strong, we followed recommendations by Certo and colleagues (2016) and examined the correlation between our independent variable of interest (competition) and the inverse Mills ratio and the pseudo-R-squared from the first-stage model. The correlation between the independent variable and the inverse Mills ratio was -0.04, and the pseudo-R-squared from the first-stage model was 0.19. Both of these suggest that the strength of our exclusion restriction is sufficient to control for selection bias (Certo et al., 2016). For our hypothesis testing, we used panel regression models. Following Papke and Wooldridge (2008), we used a clustered correlation structure (grouped by firms) and robust standard errors to correctly specify our panel data structure. We also used firm fixed effects in all the second-stage models to account for time-invariant unobservable characteristics of trial sponsors.

RESULTS

The descriptive statistics and correlations for the sample are reported in Table 1. For the trials included in our sample, it took on average 854 days (2.3 years) to report the results. Furthermore, on average, the tenor of the scientific publications linked to the trial is 56% positive.

Insert Tables 1-2 & Figures 1-3 about here

Table 2 reports our hypotheses tests. Model 1 is the first-stage model with the Heckman selection ratio and all the controls. We find that the coefficient for competition is negative and significant ($b = -0.18$, $p = 0.000$), indicating that competition decreases the likelihood of disclosure, providing support for previous studies that found competition drives firms to take a closed-door stance (Devarakonda & Reuer, 2018; James et al., 2013). Model 2 is the second stage model (using the Inverse Mills Ratio) predicting the delay in disclosure. Again, the coefficient for competition is negative and significant on the delay of disclosure ($b = -49.97$, $p = 0.005$). We draw the effect size in Figure 1. At the mean value of competition, the predicted time to report the results after the completion date is approximately 860 days, after accounting for other covariates in the model at their mean values. With a one standard deviation increase in competition, the delay reduces to about 817 days. Thus, we received support for Hypothesis 1. A coefficient to note is that of Phase 3 or 4 (late-stage trials), which is positive in Model 1 and negative in Model 2. This suggests that late-stage trials are more likely to have their results reported compared to early-stage trials, and in addition, their results will be shared more quickly (i.e., with less delay). This result further lends support to our argument for competitive signaling. Since the firms perceive late-stage trials as more appropriable, they signal more and move faster to claim disease territories.

Model 3 is the second-stage model predicting the tenor of publications. The coefficient of competition is positive and significant ($b = 0.03$, $p = 0.006$). We plot the effect size in Figure 2. At the mean value of the competition, the tenor of publications linked to the trials is 56.03%, after accounting for other covariates in the model at their mean values. With a one standard deviation increase in the perceived competition, the tenor increases to 58.65%. Thus, we received support for Hypothesis 2. Another covariate to note here is the adverse event ratio, which measures the proportion of serious events out of all the events in a reported trial. We find that its coefficient is negative and significant on the tenor of publications, confirming that a positive tenor of publications correlates with proxies of positive results for a trial (less adverse events).

In Model 4, we introduced the interaction term of reputation with competition on disclosure delay and found it to be non-significant ($b = 29.24$, $p = 0.323$). Thus, we received no support for Hypothesis 3a. In model 5, we introduced the interaction term of reputation with competition on disclosure tenor and found it to be significantly positive ($b = 0.02$, $p = 0.043$). Thus, we received support for Hypothesis 3b. We plot the interaction in Figure 3, in which we look at the disclosure of two polar groups—sponsors with an exceptional reputation for getting new drugs approved (top five percentile) and those with a little reputation (bottom five percentile)—and find that at high values of competition, highly reputed players use a more positive tenor in their publication relative to less reputed players.

Robustness and Extensions

Hazard Model. Our delay variable captures disclosure in the days since trial completion. To examine how competition influences the rate of disclosure at a particular point in time since trial completion, we recoded the variable as a hazard rate (i.e., the probability of reporting at time t), and ran a Cox regression model. The results using this approach (see Appendix D) are qualitatively in line with our main analysis.

Omitted variable bias. Although we include in all models several controls, including for therapeutic areas, some unobserved variables may still be driving both competition and reporting behaviors, such as how lucrative a sponsor might think it is to find a cure for a certain disease area. These missing variables could induce endogeneity in our analysis. To quantify the sensitivity of our analysis to potential omitted variables, we rely on the Impact Threshold for a Confounding Variable (ITCV) approach (Busenbark, Yoon, Gamache, & Withers, 2022). The ITCV describes the minimum correlation necessary between an omitted variable and the key variables of a study to create a spurious statistical inference. We find that to invalidate the inferences in our three models—predicting likelihood, speed, and tenor of disclosure—50.62%, 31.24%, and 28.36%, respectively, of the estimates, would have to be due to bias. In other words, the above percentage of cases would have to be replaced with cases for which there is an effect of zero (Xu, Frank, Maroulis, & Rosenberg, 2019).

Mechanism – Strategic signaling. To assess publication tenor in our main analysis, we consider only the publications that firms voluntarily decide to link to their ClinicalTrials.gov records. It might be the case that the firm's decision to link to a certain publication is contingent on whether the publication's results are favorable for the study investigators. If this were the case, we would find further support for our argument that the firms strategically signal on ClinicalTrials.gov and that scientific papers attached play a role in signaling. To explore this issue, we investigate whether there are any systematic differences in the language of the papers that the firms choose to share on ClinicalTrials.gov versus those they decide not to share. This requires identifying papers that could have been potentially shared, e.g., because they report the results of a ClinicalTrials.gov trial. To do so, we exploit the fact that publications linked to a trial registered on ClinicalTrials.gov should report the trial identification number (NCT ID) in their title. We searched in PubMed for the NCT IDs of all trials where the trial sponsors did not share any publication on the registry. We then calculated the tenor of the abstracts for these papers and, through t-testing, found that they were significantly lower in the positive tenor (Mean Difference = 0.13, t = 43.98) than the publications found in ClinicalTrials.gov.

DISCUSSION

We set out to investigate the competitive signaling behavior of firms when regulatory disclosure mandates shape their information environment. We examine the implicit assumption suggested in the extant literature that there is limited signaling behavior under the influence of regulatory disclosures and that firms facing competition reluctantly comply with them. Instead, we find that competition in a regulatory forum drives firms to signal through the speed and tenor of their releases. Further, firms with a high reputation seek signal congruence and are more inclined to competitive signaling.

Theoretical Contributions

Our study provides several theoretical contributions to competitive signaling and regulatory disclosure theories. First, we relax the assumption that firms only signal with voluntary information (Connelly et al., 2011; Gao et al., 2016; Guo et al., 2017; Heil & Robertson, 1991). Combining insights from the literature on the externalities of regulatory disclosure (Aghamolla & Thakor, 2022; Aragon-Correa et al., 2020; Leuz & Wysocki, 2016; Mahoney, 1995; North, 1990), we argue that regulatory mandates change the information environment of the firm and, in turn, make competition in a regulatory forum particularly salient for firms. This integration helps us test the implicit assumption held by extant literature that there is scant room for strategic maneuvers under regulatory oversight (Desai, 2016; Fabrizio & Kim, 2019; Oliver, 1991; Shi & Connelly, 2018). Thus, we show that competitive signaling behavior might stretch beyond the current bounds of voluntary disclosures where firms enjoy complete control over signals. Under regulatory constraints, firms can comply with regulators *and* signal to competitors.

While accounting and financial information disclosures are common (Cianciaruso & Sridhar, 2018; Einhorn, 2005; Mahoney, 1995), only a few industries require firms to disclose details about R&D projects to public platforms. These mandatory disclosures become especially crucial in industries such as pharmaceuticals and biotech that are knowledge-based, and disclosures can shift the competitive landscape. Across industries, innovative firms *voluntarily* share information on their inventions to seek intellectual property rights or share updates on innovative product development projects to boost stakeholder confidence and deter competition (Guo, Lev, & Zhou, 2004; Hwang, 2013; Levitas et al., 2016; Sharma & Lacey, 2004). On the surface, the *required* disclosures might seem straightforward as firms are assumed to have a restricted choice of either complying completely or risking facing non-compliance penalties (both monetary and reputational). We question this simplistic model to develop a more nuanced view of a firm's conduct facing required knowledge disclosures.

Second, by investigating the links between competition and tenor, we offer evidence of firms' linguistic games in their scientific publications to portray their studies in a flattering light to their competitors. In doing so, we contribute to competitive signaling (Guo et al., 2017; Heil & Robertson, 1991) and, more recently, rhetoric signaling theories (Gutiérrez et al., 2020;

Steigenberger & Wilhelm, 2018) in understanding competitive contexts where knowledge is embedded in language. So far, the scientific text is considered free from language games as it is meant to "report" findings to other scientists and knowledge seekers. However, with more advanced textual analysis tools, researchers have started investigating language games in scientific texts (Markowitz & Hancock, 2016; Pennebaker, 2011; Sell & Farreras, 2017). We extend this investigation to a competitive arena. When firms try to ward off competitive threats, they employ every possible signal in their arsenal, including rhetorical.

Third, we go beyond publication assessment and selective reporting bias (DeVito & Goldacre, 2019; Dwan et al., 2013; Salandra, 2018) to show that firms are employing subtler strategies to signal to competitors. Extant literature has extensively studied publication bias. Firms selectively report positive results and keep out negative findings in their scientific publications. In other words, we add to the understanding of *what* with *how* firms report their findings in knowledge texts. Arguably, employing crafty language in scientific texts is a subtler form of strategic maneuver than selective reporting in a knowledge-based industry.

Practical Contributions

This study has important implications for policymakers who passed the FDAAA 801 to increase transparency, efficiency, public trust, and enrollment in clinical trials (US Federal Government., 2007). Although researchers have studied various responses by trial sponsors (DeVito et al., 2020; Goldacre, 2013; Hsu et al., 2019; Zarin, Tse, Williams, & Carr, 2016), they have not yet considered the externalities of regulatory mandates affecting a firm's competitive environment. Our results indicate that speedy compliance with the regulatory requirement depends on the competition for the sponsors' disease conditions. Since FDAAA 801 opens a window into the advanced R&D

activities of a pharmaceutical firm to its competitors (Castellani, 2013), policymakers need to be aware of competition as an important factor shaping compliance.

Policymakers also need to be aware of competitive signaling through scientific publications. Although the positive tenor in scientific publications may be a subtle form of manipulation, its implications are not. For instance, Gigerenzer and colleagues (2007) have found extensive evidence that, often, medical doctors are "statistically illiterate" and rely on how information is framed for them by pharmaceutical firms. Because "medical journals often report evidence in nontransparent forms that suggest big benefits of featured interventions and small harms (Gigerenzer et al., 2007, p. 57)", the positive tenor in medical publications can have severe consequences. If medical researchers are susceptible to scientific literature framing, then the risk of communication with patients (Gigerenzer, 2015) and the further progress of science in disease conditions with highly biased publications are hampered. Hence, positive framing can waste substantial human and financial resources (Glasziou et al., 2014), mislead doctors, and harm patients (Goldacre, 2014).

Limitations, Future Research, and Conclusion

This study has several limitations that offer important directions for future research. One limitation is that we do not measure the consequences of firm signals on competitors and stakeholders, e.g., the response by competitors to firm signals and the effects of reporting and tenor on external stakeholders, including patients and follow-on researchers. For example, how do rival firms respond to the focal firm's signals in the regulatory platform? We directed our effort at the propensity of signaling rather than at its implications. We hope future work can delve into testing the implications of signals on the subsequent moves of firms.

Second, context plays a significant role in our study. The biopharma industry is heavily regulated, and regulatory demand for R&D disclosures is becoming increasingly prevalent. Most recently, House and Senate Democrats, on April 7, 2022, introduced legislation that mandates publicly traded drug makers to disclose research and development costs (117th US Congress, 2022). While our theory does not assume any specific information context and applies to all regulatory disclosures, we believe the tested relationships will be weaker for non-R&D disclosures. For starters, R&D proprietary information carries greater competitor threats. The uncertain nature of the product development process and innovation activity play a role in accentuating competitive signaling behavior. It would be interesting to know if the findings can be extended to other contexts with regulated disclosures (e.g., financial accounting). Finally, although we try to address endogeneity, our setting is not a controlled lab. We hope that some of the future research will be better at identifying causality.

We have elucidated information signals that firms might employ when facing intense competition and in a regulatory information environment. By integrating the literature on regulatory disclosure externalities and competitive signaling theory, we showed that firms facing intense competition move fast to signal primacy. Further, competition creates an inducement to use rhetorical signaling in their scientific publications linked to disclosures. We hope our study elucidates firms' competitive signaling when facing regulatory pressures.

REFERENCES

117th US Congress. 2022, April 7. S.4037 - 117th Congress (2021-2022): Pharmaceutical Research Transparency Act of 2022. legislation. Retrieved August 5, 2022, from https://www.congress.gov/bill/117th-congress/senate-bill/4037?s=1&r=13

Aghamolla, C., & Thakor, R. T. 2022. Do mandatory disclosure requirements for private firms increase the propensity of going public? *Journal of Accounting Research*, 60(3): 755–804.

Ahuja, G., & Yayavaram, S. 2011. PERSPECTIVE—Explaining Influence Rents: The Case for an Institutions-Based View of Strategy. *Organization Science*, 22(6): 1631–1652.

Anderson, M. L., Chiswell, K., Peterson, E. D., Tasneem, A., Topping, J., & Califf, R. M. 2015. Compliance with Results Reporting at ClinicalTrials.gov. *New England Journal of Medicine*, 372(11): 1031–1039.

Aragon-Correa, J. A., Marcus, A. A., & Vogel, D. 2020. The effects of mandatory and voluntary regulatory pressures on firms' environmental strategies: A review and recommendations for future research. *Academy of Management Annals*, 14(1): 339–365.

Azoulay, P. 2002. Do Pharmaceutical Sales Respond to Scientific Evidence? *Journal of Economics & Management Strategy*, 11(4): 551–594.

Baker, S., & Mezzetti, C. 2005. Disclosure as a Strategy in the Patent Race. *The Journal of Law and Economics*, 48(1): 173–194.

Ball, G. P., Shah, R., & Wowak, K. D. 2018. Product competition, managerial discretion, and manufacturing recalls in the U.S. pharmaceutical industry. *Journal of Operations Management*, 58–59(1): 59–72.

Bapna, S. 2019. Complementarity of Signals in Early-Stage Equity Investment Decisions: Evidence from a Randomized Field Experiment. *Management Science*, 65(2): 933–952.

Bayus, B. L., Jain, S., & Rao, A. G. 2001. Truth or Consequences: An Analysis of Vaporware and New Product Announcements. *Journal of Marketing Research*, 38(1): 3–13.

Bergh, D. D., Connelly, B. L., Ketchen, D. J., & Shannon, L. M. 2014. Signalling Theory and Equilibrium in Strategic Management Research: An Assessment and a Research Agenda. *Journal of Management Studies*, 51(8): 1334–1360.

Blackburne, T. P., & Quinn, P. J. 2020. Disclosure speed: Evidence from nonpublic SEC investigations. *The Accounting Review*.

Booth, J. A. 2020, August 21. Using Clinicaltrials.gov for Competitor Analysis. *Cipher*. Retrieved from https://www.cipher-sys.com/blog/using-clinicaltrialsgov-for-competitor-analysis

Busenbark, J. R., Yoon, H. (Elle), Gamache, D. L., & Withers, M. C. 2022. Omitted Variable Bias: Examining Management Research With the Impact Threshold of a Confounding Variable (ITCV). *Journal of Management*, 48(1): 17–48.

Cade, N. L., Kaplan, S. E., & Loftus, S. 2021. Changing the blame game: Does the presence of a pay ratio disclosure impact nonprofessional investors' reactions to CEOs' internal attributions for poor firm performance? *Journal of Management Accounting Research*.

Caldwell, P. H. Y., Hamilton, S., Tan, A., & Craig, J. C. 2010. Strategies for Increasing Recruitment to Randomised Controlled Trials: Systematic Review. *PLOS Medicine*, 7(11): e1000368.

Caner, T., Bruyaka, O., & Prescott, J. E. 2018. Flow Signals: Evidence from Patent and Alliance Portfolios in the US Biopharmaceutical Industry. *Journal of Management Studies*, 55(2): 232–264.

Carpenter, G. S., & Nakamoto, K. 1989. Consumer preference formation and pioneering advantage. *Journal of Marketing research*, 26(3): 285–298.

Castellani, J. 2013. Are clinical trial data shared sufficiently today? Yes. *BMJ*, 347: f1881.

Certo, S. T., Busenbark, J. R., Woo, H., & Semadeni, M. 2016. Sample selection bias and Heckman models in strategic management research. *Strategic Management Journal*, 37(13): 2639–2657.

Cha, M., & Yu, F. 2014. *Pharma's first-to-market advantage.* McKinsey & Company. Retrieved from https://www.mckinsey.com/industries/pharmaceuticals-and-medical-products/our-insights/pharmas-first-to-market-advantage

Chen, M., Su, K.-H., & Tsai, W. 2007. Competitive Tension: The Awareness-Motivation-Capability Perspective. *Academy of Management Journal*, 50(1): 101–118.

Chen, M.-J. 1996. Competitor Analysis and Interfirm Rivalry: Toward A Theoretical Integration. *Academy of Management Review*, 21(1): 100–134.

Chen, M.-J., & Hambrick, D. C. 1995. Speed, Stealth, and Selective Attack: How Small Firms Differ From Large Firms in Competitive Behavior. *Academy of Management Journal*, 38(2): 453–482.

Chen, M.-J., & Miller, D. 2012. Competitive Dynamics: Themes, Trends, and a Prospective Research Platform. *Academy of Management Annals*, 6(1): 135–210.

Choi, Y., Chiu, C. Y.-I., & Sontag, D. 2016. Learning low-dimensional representations of medical concepts. *AMIA Summits on Translational Science Proceedings*, 2016: 41.

Cianciaruso, D., & Sridhar, S. S. 2018. Mandatory and Voluntary Disclosures: Dynamic Interactions. *Journal of Accounting Research*, 56(4): 1253–1283.

Clark, B. H., & Montgomery, D. B. 1998. Deterrence, Reputations, and Competitive Cognition. *Management Science*, 44(1): 62–82.

Clarkson, G., & Toh, P. K. 2010. 'Keep out' signs: the role of deterrence in the competition for resources. *Strategic Management Journal*, 31(11): 1202–1225.

Cohen, W. M., Goto, A., Nagata, A., Nelson, R. R., & Walsh, J. P. 2002. R&D spillovers, patents and the incentives to innovate in Japan and the United States. *Research Policy*, 31(8–9): 1349–1367.

Connelly, B. L., Certo, S. T., Ireland, R. D., & Reutzel, C. R. 2011. Signaling Theory: A Review and Assessment. *Journal of Management*, 37(1): 39–67.

Dahm, M., Gonzalez, P., & Porteiro, N. 2009. Trials, tricks and transparency: how disclosure rules affect clinical knowledge. *J Health Econ*, 28(6): 1141–53.

De Chernatony, L., Daniels, K., & Johnson, G. 1993. A cognitive perspective on managers' perceptions of competition. *Journal of Marketing Management*, 9(4): 373–381.

Desai, V. M. 2016. Under the Radar: Regulatory Collaborations and their Selective Use to Facilitate Organizational Compliance. *Academy of Management Journal*, 59(2): 636–657.

Devarakonda, S. V., & Reuer, J. J. 2018. Knowledge sharing and safeguarding in R&D collaborations: The role of steering committees in biotechnology alliances. *Strategic Management Journal*, 39(7): 1912–1934.

DeVito, N. J., Bacon, S., & Goldacre, B. 2020. Compliance with legal requirement to report clinical trial results on ClinicalTrials.gov: a cohort study. *The Lancet*, 395(10221): 361–369.

DeVito, N. J., & Goldacre, B. 2019. Catalogue of bias: publication bias. *BMJ Evidence-Based Medicine*, 24(2): 53.

Dimov, D., Shepherd, D. A., & Sutcliffe, K. M. 2007. Requisite expertise, firm reputation, and status in venture capital investment allocation decisions. *Journal of Business Venturing*, 22(4): 481–502.
Downing, S. T., Kang, J.-S., & Markman, G. D. 2019. What you don't see can hurt you: Awareness cues to profile indirect competitors. *Academy of Management Journal*, 62(6): 1872–1900.
Dranove, D., & Gandal, N. 2003. The Dvd-vs.-Divx Standard War: Empirical Evidence of Network Effects and Preannouncement Effects. *Journal of Economics & Management Strategy*, 12(3): 363–386.
Drover, W., Wood, M. S., & Corbett, A. C. 2018. Toward a Cognitive View of Signalling Theory: Individual Attention and Signal Set Interpretation. *Journal of Management Studies*, 55(2): 209–231.
Dwan, K., Gamble, C., Williamson, P. R., & Kirkham, J. J. 2013. Systematic Review of the Empirical Evidence of Study Publication Bias and Outcome Reporting Bias — An Updated Review. *PLOS ONE*, 8(7): e66844.
Edelman, L. B., Uggen, C., & Erlanger, H. S. 1999. The Endogeneity of Legal Regulation: Grievance Procedures as Rational Myth. *American Journal of Sociology*, 105(2): 406–54.
Einhorn, E. T. I. 2005. The Nature of the Interaction between Mandatory and Voluntary Disclosures. *Journal of Accounting Research*, 43(4): 593–621.
Fabrizio, K. R., & Kim, E.-H. 2019. Reluctant Disclosure and Transparency: Evidence from Environmental Disclosures. *Organization Science*, 30(6): 1207–1231.
FDA. 2021. Drugs@FDA Glossary. Retrieved June 23, 2021, from https://www.accessdata.fda.gov/scripts/cder/daf/index.cfm?event=glossary.page
Federal Register. 2016. Retrieved from https://www.govinfo.gov/content/pkg/FR-2016-09-21/pdf/2016-22129.pdf
Gao, H., Yu, T., & Cannella, A. A. 2016. The Use of Public Language in Strategy: A Multidisciplinary Review and Research Agenda. *Journal of Management*, 42(1): 21–54.
Gigerenzer, G. 2015. *Risk Savvy: How to Make Good Decisions*. Penguin.
Gigerenzer, G., Gaissmaier, W., Kurz-Milcke, E., Schwartz, L. M., & Woloshin, S. 2007. Helping Doctors and Patients Make Sense of Health Statistics. *Psychological Science in the Public Interest*, 8(2): 53–96.
Glasziou, P., Altman, D. G., Bossuyt, P., Boutron, I., Clarke, M., Julious, S., Michie, S., Moher, D., & Wager, E. 2014. Reducing waste from incomplete or unusable reports of biomedical research. *The Lancet*, 383(9913): 267–276.
Goldacre, B. 2013. Are clinical trial data shared sufficiently today? No. *BMJ*, 347: f1880.
Goldacre, B. 2014. *Bad Pharma: How Drug Companies Mislead Doctors and Harm Patients*. Macmillan.
Grabowski, H. G., & Vernon, J. M. 1992. Brand Loyalty, Entry, and Price Competition in Pharmaceuticals after the 1984 Drug Act. *The Journal of Law and Economics*, 35(2): 331–350.
Graffin, S. D., Haleblian, J., & Kiley, J. T. 2016. Ready, AIM, Acquire: Impression Offsetting and Acquisitions. *Academy of Management Journal*, 59(1): 232–252.
Grant, R. M. 1996. Toward a knowledge-based theory of the firm. *Strategic Management Journal*, 17(S2): 109–122.
Grossman, S. J. 1981. The Informational Role of Warranties and Private Disclosure about Product Quality. *The Journal of Law & Economics*, 24(3): 461–483.

Guo, R.-J., Lev, B., & Zhou, N. A. N. 2004. Competitive Costs of Disclosure by Biotech IPOs. *Journal of Accounting Research*, 42(2): 319–355.

Guo, W., Sengul, M., & Yu, T. 2020. Rivals' Negative Earnings Surprises, Language Signals, and Firms' Competitive Actions. *Academy of Management Journal*, 63(3): 637–659.

Guo, W., Yu, T., & Gimeno, J. 2017. Language and Competition: Communication Vagueness, Interpretation Difficulties, and Market Entry. *Academy of Management Journal*, 60(6): 2073–2098.

Gutiérrez, M., Papiashvili, N., Tribó, J. A., & Vazquez, A. B. 2020. Managerial incentives for attracting attention. *European Financial Management*, 26(4): 896–937.

Haan, M. A. 2003. Vaporware as a Means of Entry Deterrence. *The Journal of Industrial Economics*, 51(3): 345–358.

Healy, P. M., & Palepu, K. G. 2001. Information asymmetry, corporate disclosure, and the capital markets: A review of the empirical disclosure literature. *Journal of Accounting and Economics*, 31(1): 405–440.

Heckman, J. J. 1979. Sample Selection Bias as a Specification Error. *Econometrica*, 47(1): 153–161.

Heil, O., & Robertson, T. S. 1991. Toward a theory of competitive market signaling: A research agenda. *Strategic Management Journal*, 12(6): 403–418.

Hsu, P.-H., Lee, K., Moon, K., & Oh, S. 2019. Information Transparency in Drug Development: Evidence from Mandatory Disclosure of Clinical Trials. *SSRN Electronic Journal*.

Hwang, T. J. 2013. Stock market returns and clinical trial results of investigational compounds: an event study analysis of large biopharmaceutical companies. *PLoS One*, 8(8): e71966.

James, S. D., Leiblein, M. J., & Lu, S. 2013. How Firms Capture Value From Their Innovations. *Journal of Management*, 39(5): 1123–1155.

Jansen, J. 2010. Strategic Information Disclosure and Competition for an Imperfectly Protected Innovation*. *The Journal of Industrial Economics*, 58(2): 349–372.

Jayaraman, S., & Wu, J. S. 2019. Is silence golden? Real effects of mandatory disclosure. *The Review of Financial Studies*, 32(6): 2225–2259.

Kajüter, P., Klassmann, F., & Nienhaus, M. 2019. The effect of mandatory quarterly reporting on firm value. *The Accounting Review*, 94(3): 251–277.

Lanjouw, J. O., & Schankerman, M. 2001. Characteristics of Patent Litigation: A Window on Competition. *The RAND Journal of Economics*, 32(1): 129–151.

Leuz, C., & Wysocki, P. D. 2016. The Economics of Disclosure and Financial Reporting Regulation: Evidence and Suggestions for Future Research. *Journal of Accounting Research*, 54(2): 525–622.

Lev, B. 1992. Information Disclosure Strategy. *California Management Review*, 34(4): 9–32.

Levitas, E., McFadyen, M. A., & Ahsan, M. 2016. The Determinants of Voluntary Disclosures in R&D Alliances. *Academy of Management Proceedings*, 2016(1): 11446.

Liberti, J. M., & Petersen, M. A. 2018. Information: Hard and Soft. *The Review of Corporate Finance Studies*, 8(1): 1–41.

Lieberman, M. B., & Montgomery, D. B. 1988. First-mover advantages. *Strategic Management Journal*, 9(S1): 41–58.

Litz, R. A., & Pollack, J. M. 2015. Interfirm Rivalry between Small Hardware Stores and "Big Box" Retailers: Market Commonality and Product Mix Similarity as Antecedents to Competitive Response. *Journal of Small Business Management*, 53(2): 436–449.

Livengood, R. S., & Reger, R. K. 2010. That's Our Turf! Identity Domains and Competitive Dynamics. *Academy of Management Review*, 35(1): 48–66.

Love, E. G., & Kraatz, M. 2009. Character, conformity, or the bottom line? How and why downsizing affected corporate reputation. *Academy of management Journal*, 52(2): 314–335.

Mahoney, P. G. 1995. Mandatory Disclosure as a Solution to Agency Problems. *The University of Chicago Law Review*, 62(3): 1047.

Markowitz, D. M., & Hancock, J. T. 2016. Linguistic Obfuscation in Fraudulent Science. *Journal of Language and Social Psychology*, 35(4): 435–445.

Marquis, C., Toffel, M. W., & Zhou, Y. 2016. Scrutiny, norms, and selective disclosure: A global study of greenwashing. *Organization Science*, 27(2): 483–504.

Mc Namara, P., & Baden-Fuller, C. 2007. Shareholder returns and the exploration–exploitation dilemma: R&D announcements by biotechnology firms. *Research Policy*, Biotechnology: Its origins, organization, and outputs, 36(4): 548–565.

McCann, B. T., & Bahl, M. 2017. The influence of competition from informal firms on new product development. *Strategic Management Journal*, 38(7): 1518–1535.

Milgrom, P. R. 1981. Good News and Bad News: Representation Theorems and Applications. *The Bell Journal of Economics*, 12(2): 380–391.

Nadkarni, S., Pan, L., & Chen, T. 2019. Only Timeline Will Tell: Temporal Framing of Competitive Announcements and Rivals' Responses. *Academy of Management Journal*, 62(1): 117–143.

Ndofor, H. A., & Levitas, E. 2004. Signaling the Strategic Value of Knowledge. *Journal of Management*, 30(5): 685–702.

Nelson, A. J. 2016. How to Share "A Really Good Secret": Managing Sharing/Secrecy Tensions Around Scientific Knowledge Disclosure. *Organization Science*, 27(2): 265–285.

North, D. C. 1990. *Institutions, Institutional Change and Economic Performance*. Political Economy of Institutions and Decisions. Cambridge: Cambridge University Press.

Oliver, C. 1991. Strategic responses to institutional processes. *Academy of Management Review*, 16(1): 145–179.

Papke, L. E., & Wooldridge, J. M. 2008. Panel data methods for fractional response variables with an application to test pass rates. *Journal of Econometrics*, 145(1): 121–133.

Pennebaker, J. W. 2011. *The secret life of pronouns: What our words say about us*. The secret life of pronouns: What our words say about us. New York, NY: Bloomsbury Press/Bloomsbury Publishing.

Pennebaker, J. W., Booth, R. J., Boyd, R. L., & Francis, M. E. 2015. Linguistic inquiry and word count: LIWC2015. Austin, TX: Pennebaker Conglomerates.

Pfarrer, M. D., Pollock, T. G., & Rindova, V. P. 2010. A Tale of Two Assets: The Effects of Firm Reputation and Celebrity on Earnings Surprises and Investors' Reactions. *Academy of Management Journal*, 53(5): 1131–1152.

Polidoro, F. Jr., & Theeke, M. 2012. Getting Competition Down to a Science: The Effects of Technological Competition on Firms' Scientific Publications. *Organization Science*, 23(4): 1135–1153.

Porac, J. F., Thomas, H., & Baden-Fuller, C. 1989. Competitive Groups as Cognitive Communities: The Case of Scottish Knitwear Manufacturers*. *Journal of Management Studies*, 26(4): 397–416.

Quigley, T. J., Hubbard, T. D., Ward, A., & Graffin, S. D. 2020. Unintended Consequences: Information Releases and CEO Stock Option Grants. *Academy of Management Journal*, 63(1): 155–180.

Raaijmakers, A. G. M., Vermeulen, P. A. M., Meeus, M. T. H., & Zietsma, C. 2015. I Need Time! Exploring Pathways to Compliance under Institutional Complexity. *Academy of Management Journal*, 58(1): 85–110.

Roychowdhury, S., Shroff, N., & Verdi, R. S. 2019. The effects of financial reporting and disclosure on corporate investment: A review. *Journal of Accounting and Economics*, 68(2): 101246.

Salandra, R. 2018. Knowledge dissemination in clinical trials: Exploring influences of institutional support and type of innovation on selective reporting. *Research Policy*, 47(7): 1215–1228.

Scherer, F. M. 2010. Chapter 12 - Pharmaceutical Innovation. In B. H. Hall & N. Rosenberg (Eds.), *Handbook of the Economics of Innovation*, Handbook of The Economics of Innovation, Vol. 1 (Vol. 1, pp. 539–574). North-Holland.

Sell, J., & Farreras, I. G. 2017. LIWC-ing at a Century of Introductory College Textbooks: Have the Sentiments Changed? *Procedia Computer Science*, 118: 108–112.

Sharma, A., & Lacey, N. 2004. Linking Product Development Outcomes to Market Valuation of the Firm: The Case of the U.S. Pharmaceutical Industry*. *Journal of Product Innovation Management*, 21(5): 297–308.

Sheth, J. D. 2021. Disclosure of information under competition: An experimental study. *Games and Economic Behavior*, 129: 158–180.

Shi, W., & Connelly, B. L. 2018. Is regulatory adoption ceremonial? Evidence from lead director appointments. *Strategic Management Journal*, 39(8): 2386–2413.

Si, Y., Wang, J., Xu, H., & Roberts, K. 2019. Enhancing clinical concept extraction with contextual embeddings. *Journal of the American Medical Informatics Association*, 26(11): 1297–1304.

Steigenberger, N., & Wilhelm, H. 2018. Extending Signaling Theory to Rhetorical Signals: Evidence from Crowdfunding. *Organization Science*, 29(3): 529–546.

Stein, J. C. 1989. Cheap Talk and the Fed: A Theory of Imprecise Policy Announcements. *The American Economic Review*, 79(1): 32–42.

Stern, I., Dukerich, J. M., & Zajac, E. 2014. Unmixed signals: How reputation and status affect alliance formation. *Strategic Management Journal*, 35(4): 512–531.

Stern, I., & James, S. D. 2016. Whom are you promoting? Positive voluntary public disclosures and executive turnover: Positive Voluntary Public Disclosures and Executive Turnover. *Strategic Management Journal*, 37(7): 1413–1430.

Tausczik, Y. R., & Pennebaker, J. W. 2010. The Psychological Meaning of Words: LIWC and Computerized Text Analysis Methods. *Journal of Language and Social Psychology*, 29(1): 24–54.

Tian, X. 2015. Does Real-Time Reporting Deter Strategic Disclosures by Management? The Impact of Real-Time Reporting and Event Controllability on Disclosure Bunching. *The Accounting Review*, 90(5): 2107–2139.

Toh, P. K., & Miller, C. D. 2017. Pawn to Save a Chariot, or Drawbridge Into the Fort? Firms' Disclosure During Standard Setting and Complementary Technologies Within Ecosystems: Pawn to Save a Chariot, or Drawbridge Into the Fort? *Strategic Management Journal*, 38(11): 2213–2236.

Trapido, D. 2013. Dual Signals: How Competition Makes or Breaks Interfirm Social Ties. *Organization Science*, 24(2): 498–512.

Tse, T., Williams, R. J., & Zarin, D. A. 2009. Reporting "basic results" in ClinicalTrials.gov. *Chest*, 136(1): 295–303.

US Federal Government. 2007. *FDAAA 801*. Retrieved from https://www.govinfo.gov/content/pkg/PLAW-110publ85/pdf/PLAW-110publ85.pdf#page=82

Vine, L. D., Zuccon, G., Koopman, B., Sitbon, L., & Bruza, P. 2014. Medical Semantic Similarity with a Neural Language Model (pp. 1819–1822). Shanghai, China: Association for Computing Machinery.

Westphal, J. D., & Bednar, M. K. 2008. The Pacification of Institutional Investors. *Administrative Science Quarterly*, 53(1): 29–72.

Winter, S. G. 2006. The logic of appropriability: From Schumpeter to Arrow to Teece. *Research Policy*, Special issue commemorating the 20th Anniversary of David Teece's article, "Profiting from Innovation", in Research Policy, 35(8): 1100–1106.

Wu, D. T. Y., Hanauer, D. A., Mei, Q., Clark, P. M., An, L. C., Proulx, J., Zeng, Q. T., Vydiswaran, V. G. V., Collins-Thompson, K., & Zheng, K. 2016. Assessing the readability of ClinicalTrials.gov. *Journal of the American Medical Informatics Association: JAMIA*, 23(2): 269–275.

Xu, R., Frank, K. A., Maroulis, S. J., & Rosenberg, J. M. 2019. konfound: Command to quantify robustness of causal inferences. *The Stata Journal*, 19(3): 523–550.

Zarin, D. A., & Tse, T. 2008. Medicine. Moving toward transparency of clinical trials. *Science*, 319(5868): 1340–1342.

Zarin, D. A., Tse, T., Williams, R. J., & Carr, S. 2016. Trial Reporting in ClinicalTrials.gov - The Final Rule. *N Engl J Med*, 375(20): 1998–2004.

Zavyalova, A., Pfarrer, M. D., Reger, R. K., & Hubbard, T. D. 2016. Reputation as a Benefit and a Burden? How Stakeholders' Organizational Identification Affects the Role of Reputation Following a Negative Event. *Academy of Management Journal*, 59(1): 253–276.

Zhou, Y. 2022. Proprietary Costs and Corporate Lobbying Against Changes in Mandatory Disclosure. *Management Science*.

FOOTNOTES

1. The law requires all interventional studies (or clinical trials) of new drugs, biologics, or devices (whether or not approved for marketing) in phases two through four, with at least one US site, to be reported. The law was further expanded and clarified with the Final Rule, effective as of January 18, 2017. Retrieved from – https://prsinfo.clinicaltrials.gov/.
2. The registration for "applicable clinical trials" (ACTs) requires information on the timeline and overall study plan, including specific details like the targeted disease or condition, the interventions tested, recruitment criteria, and location of trial sites. For detailed data elements reported as part of registration, refer to https://prsinfo.clinicaltrials.gov/definitions.html.
3. The ClinicalTrials.gov's results-quality-control-review-guide specifies general criteria: "Written results or conclusions are not presented in any free-text field as the only means of reporting data." Retrieved from – https://prsinfo.clinicaltrials.gov/ResultsDetailedReviewItems.pdf; page 3.
4. These disclosures are aimed at "helping patients find trials for which they might be eligible, enhance the design of clinical trials and preventing duplication of unsuccessful or unsafe trials, improving the evidence base that informs clinical care, increasing the efficiency of drug and device development processes, improving clinical research practice, and building public trust in clinical research" (*Federal Register*, 2016, p. 64982).
5. In the interest of parsimony, we have not included the trial details as an exhibit, but they can be accessed here – https://clinicaltrials.gov/ct2/show/study/NCT01999322?term=NCT01999322&draw=2&rank=1.
6. We followed the observational medical studies for this identification process (Anderson et al., 2015).
7. Retrieved from – https://pubmed.ncbi.nlm.nih.gov/. We used only abstract text when the full-text PDFs were not extractable – about 14% of the articles.
8. In robustness checks, we tested our models only on the trials where the delay was more than one year (> 365 days) – about 95% of our sample. The results were largely unchanged.
9. The competing projects need not be ACTs (applicable clinical trials) or have the same phase. They only need matching disease conditions and an overlapping trial period with the focal trial to be counted as competing studies.
10. The annual FDA Orange Book contains all approved pharmaceutical products for sale in the US, categorized by whether the drug gained approval using the NDA or ANDA approval pathway. However, if a drug is discontinued, the FDA does not stipulate the year in which the drug moved to a discontinued status in the most recent Orange Book. To determine the exact year a drug was discontinued, which is necessary to obtain a precise measure of product competition for each firm-year, we needed to acquire the history of Orange Books published. For information and the history of the Orange Books, we consult https://www.fda.gov/drugs/drug-approvals-and-databases/additionsdeletions-prescription-and-otc-drug-product-lists.
11. Detailed steps of the methodology we followed are noted here – https://towardsdatascience.com/use-embeddings-to-predict-therapeutic-area-of-clinical-studies-654af661b949.

TABLES AND FIGURES

Table 1: Summary Statistics and Correlations

	Mean	SD	Min.	Max.	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
(1) Disclosure Delay	0.67	0.47	0	4229	1							
(2) Disclosure Tenor	853.58	671.6	0	1	0.02	1						
(3) Competition	0.56	0.13	0	4172	-0.07	0.07	1					
(4) Reputation	536.69	538.88	0	11	-0.16	0	-0.08	1				
(5) Previous Firm Trials	5.12	2.14	0	3217	-0.04	0	0.07	0.02	1			
(6) Phase3 or 4	0.14	0.37	0	1	-0.3	0	0.14	0.12	0.01	1		
(7) Product Competition	0.5	0.5	0	1	-0.02	0.17	0.04	0.09	-0.1	0.04	1	
(8) Adverse Events Ratio	0.14	0.19	0	1	-0.09	-0.06	0.2	-0.12	-0.16	-0.21	-0.05	1
(9) No. of Sponsors	1.15	0.46	1	12	0.04	0	-0.05	-0.06	0.07	-0.07	-0.09	-0.02
(10) No of Facilities	54.67	90.27	1	1621	-0.14	-0.13	0.31	-0.08	0.05	0.25	-0.12	0.14
(11) Requested Extension	0.41	0.49	0	1	0.45	-0.01	-0.01	-0.13	0.06	-0.09	-0.17	-0.14
(12) Number of Subjects (log)	5.23	1.35	0.69	10.69	-0.11	-0.16	0.28	-0.07	0.13	0.43	-0.06	-0.11
(13) Firm Size	8.69	3.04	-2.15	12.27	-0.07	-0.04	0.13	0.02	0.61	0.02	0.01	-0.09
(14) R&D Intensity	1593.16	7925.06	-62.86	210277.02	-0.02	-0.04	-0.09	-0.03	-0.1	0.07	-0.12	-0.03
(15) Firm Slack	3.54	4.54	0.02	56.99	0.06	-0.02	-0.1	-0.04	-0.41	-0.12	-0.11	0.22
(16) Heckman Selection Ratio	0.16	0.26	0	3	0.02	0.08	0.37	-0.1	-0.2	0.23	-0.03	-0.09

	Mean	SD	Min.	Max.	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
(8) Adverse Events Ratio	0.14	0.19	0	1	1								
(9) No. of Sponsors	1.15	0.46	1	12	-0.02	1							
(10) No of Facilities	54.67	90.27	1	1621	0.14	0	1						
(11) Requested Extension	0.41	0.49	0	1	-0.14	0.03	0.04	1					
(12) Number of Subjects (log)	5.23	1.35	0.69	10.69	-0.11	0.01	0.62	0.09	1				
(13) Firm Size	8.69	3.04	-2.15	12.27	-0.09	0.06	0.19	0.11	0.26	1			
(14) R&D Intensity	1593.16	7925.06	-62.86	210277.02	-0.03	-0.04	-0.04	-0.01	-0.05	-0.32	1		
(15) Firm Slack	3.54	4.54	0.02	56.99	0.22	-0.02	-0.05	0.02	-0.1	-0.54	0.11	1	
(16) Heckman Selection	0.16	0.26	0	3	-0.09	-0.18	0.03	0.06	0.08	-0.41	0.19	0.12	1

Notes. Correlation coefficients with an absolute value larger than .03 are significant at the .05 level.

Table 2: Disclosure Delay and Tenor - Heckman Selection Regressions

	(1) Likelihood	(2) Delay	(3) Tenor	(4) Delay	(5) Tenor
Heckman Selection Ratio	0.402				
	(0.012)				
IMillsR		51.612	-0.099	62.140	-0.099
		(0.718)	(0.286)	(0.666)	(0.283)
Adverse Events Ratio		-208.528	-0.133	-210.897	-0.134
		(0.003)	(0.019)	(0.003)	(0.018)
Phase 3 or 4	0.545	-231.982	-0.003	-228.479	-0.003
	(0.000)	(0.000)	(0.864)	(0.000)	(0.888)
No of Sponsors	0.027	25.745	0.009	25.715	0.010
	(0.784)	(0.441)	(0.744)	(0.447)	(0.714)
No of Facilities	-0.000	-0.200	-0.000	-0.192	0.000
	(0.386)	(0.058)	(0.933)	(0.056)	(0.975)
Requested Extension	-1.036	595.001	0.049	588.755	0.047
	(0.000)	(0.000)	(0.172)	(0.000)	(0.179)
Number of Subjects (log)	0.093	0.655	-0.023	-0.199	-0.024
	(0.073)	(0.955)	(0.012)	(0.987)	(0.009)
Firm Slack	0.006	7.834	0.002	7.841	0.002
	(0.558)	(0.215)	(0.625)	(0.215)	(0.673)
Firm Size	0.111	-72.094	0.037	-71.769	0.039
	(0.001)	(0.043)	(0.371)	(0.047)	(0.332)
Product Competition	0.498	-179.287	-0.030	-192.848	-0.053
	(0.320)	(0.578)	(0.833)	(0.550)	(0.711)
R&D Intensity	-0.000	0.006	0.000	0.006	0.000
	(0.629)	(0.232)	(0.099)	(0.236)	(0.086)
Previous Firm Trials	0.000	0.001	0.000	-0.000	0.000
	(0.469)	(0.996)	(0.492)	(1.000)	(0.440)
Reputation	0.077	-56.716	-0.012	-50.713	-0.009
	(0.135)	(0.005)	(0.288)	(0.000)	(0.411)
Competition (standardized)	-0.182	-49.969	0.030	-58.801	0.028
	(0.000)	(0.005)	(0.006)	(0.004)	(0.009)
Reputation x Competition				29.243	0.015
				(0.323)	(0.043)
Constant	-3.022	2,155.719	0.532	2,126.121	0.520
	(0.000)	(0.000)	(0.160)	(0.000)	(0.158)
Observations (Number of Trials)	4,938	3,261	584	3,261	584
Log Pseudo/R-squared	-2067.87	0.401	0.149	0.402	0.151
Number of Firms	396	209	63	209	63
Therapeutic Area FE	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes
Robust SE	Yes	Yes	Yes	Yes	Yes

Notes. Robust *p* values in parentheses. Two-tailed tests.

Figure 1

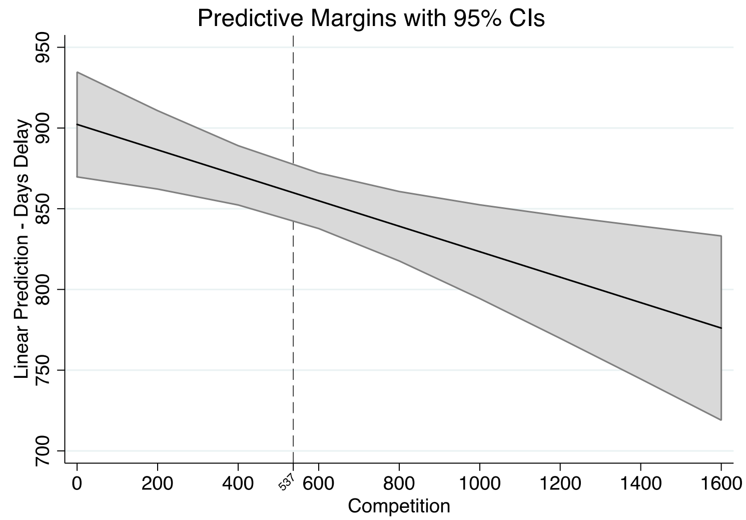


Figure 2

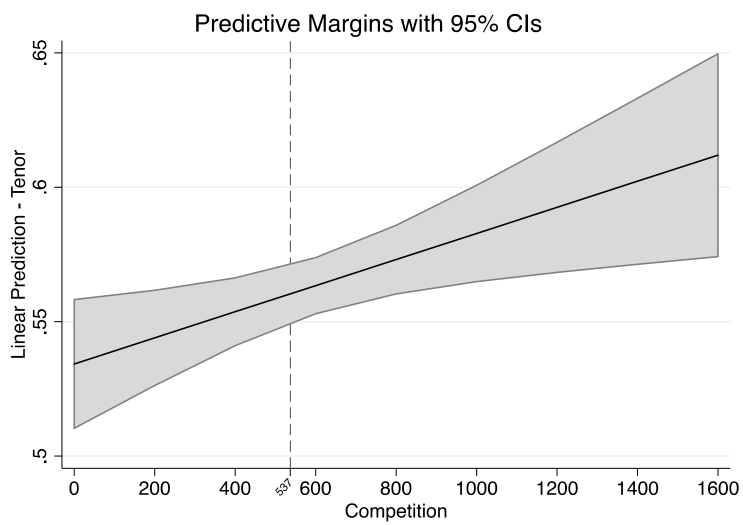
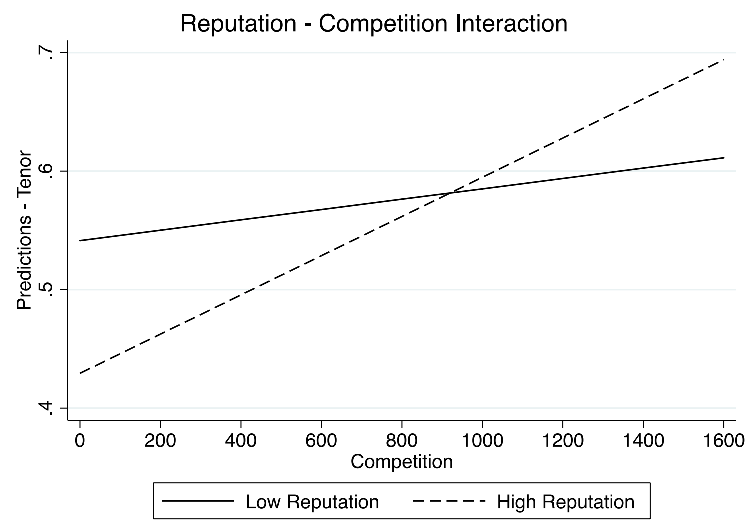
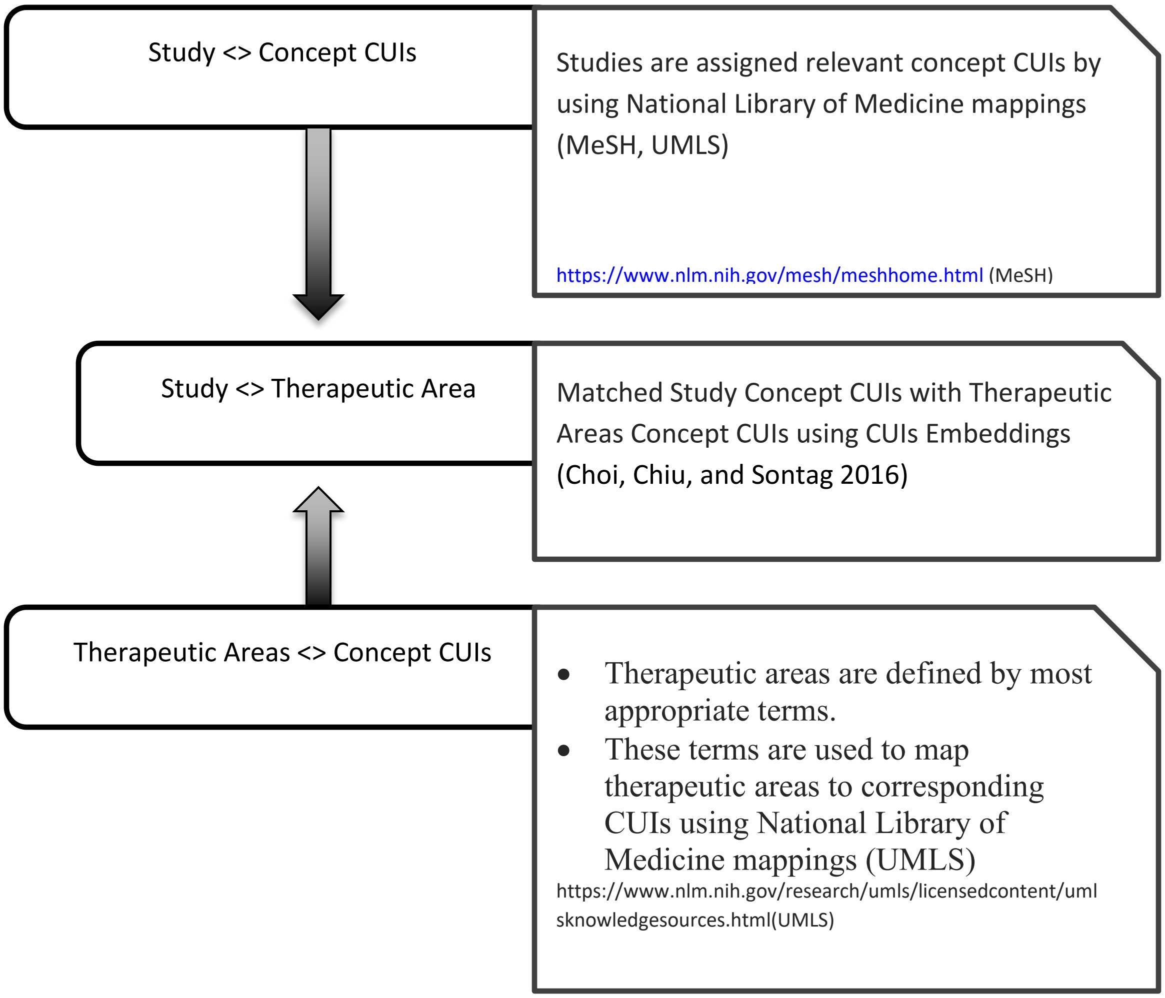


Figure 3



Appendix A: Algorithm used to assign Therapeutic Areas to Clinical Trials and Data Sources.



Appendix B: Therapeutic Areas and the associated terms.

	Therapeutic Area	Associated terms
1	Cardiology/Vascular Diseases	‘cardiolog’, ‘cardiovascul’.
2	Dental and Oral Health	‘tooth cavit’, ‘caries’, ‘cavities’ Excluded Terms-‘peccaries’, ‘cotyloid cavities’, ‘nasal cavities’, ‘cavities glenoid’, ‘cavities paranasal’, ‘pleural cavities’, ‘cavities, pleural’, ‘pericardial cavities’, ‘cavities pelvic’, ‘cavities uterine’, ‘abdominal cavities', ‘cavities tympanic’, ‘body cavities’.
3	Dermatology	‘dermatol’.
4	Devices	‘device’.
5	Disorders of Environmental Origin	‘environmental’, ‘environments’, ‘pollut’.
6	Endocrinology	‘endocrinol’.
7	Gastroenterology	‘gastroentero’.
8	Genetic Disease	‘geneti, diseas'.
9	Hematology	‘hematol' . Excluded terms-‘non-hemato’.
10	Hepatology	‘hepatitis’.
11	Immunology	‘Immunolog’.
12	Infections and Infectious Diseases	‘infectious, infected, ‘infection’, ‘disease’.
13	Internal Medicine	‘intern’, ‘medicin’.
14	Musculoskeletal	‘musculoskelet’.
15	Nephrology	‘nephrolog’.
16	Neurology	‘neurolog’.
17	Nutrition and Weight Loss	‘nutrition’, ‘body weight’, ‘weight reduc’, ‘weight gain’, ‘overweight’.
18	Obstetrics/Gynecology	‘obstetri’, ‘gynecol’.
19	Occupational Diseases	‘occupational disease’.
20	Oncology	‘oncolog’, ‘cancer’.
21	Ophthalmology	‘ophthalmol’, ‘eye’.
22	Orthopedics/Orthopedic Surgery	‘orthopedi’.
23	Parasitic Diseases	‘parasit’, ‘disease’.
24	Pediatrics/Neonatology	‘pediatr’, ‘neonat’.
25	Pharmacology/Toxicology	‘pharmacol’, ‘toxicol’.
26	Psychiatry/Psychology	‘psychiatr’, ‘psycholog’.
27	Pulmonary/Respiratory Diseases	‘pulmonar’, ‘respirat,diseas’.
28	Rheumatology	‘rheumat’.
29	Sleep	‘sleep’.
30	Symptoms and General Pathology	‘general manifestation of disorders’.
31	Trauma	‘traumas’.
32	Urology	‘urology’.

33	Vaccines	'vaccine'.

Appendix C: Naïve Models.

	(1) Likelihood	(2) Delay	(3) Tenor	(4) Delay	(5) Tenor
Adverse Events Ratio		-208.378	-0.131	-210.678	-0.132
		(0.003)	(0.018)	(0.003)	(0.017)
Phase 3 or 4	0.566	-245.479	0.023	-244.741	0.024
	(0.000)	(0.000)	(0.231)	(0.000)	(0.224)
No of Sponsors	0.025	24.952	0.009	24.763	0.010
	(0.803)	(0.464)	(0.730)	(0.472)	(0.700)
No of Facilities	-0.000	-0.189	-0.000	-0.179	-0.000
	(0.347)	(0.074)	(0.690)	(0.073)	(0.780)
Requested Extension	-1.036	620.298	0.004	619.229	0.003
	(0.000)	(0.000)	(0.801)	(0.000)	(0.859)
Number of Subjects	0.094	-1.481	-0.020	-2.750	-0.021
	(0.068)	(0.884)	(0.022)	(0.790)	(0.018)
Firm Slack	0.007	7.567	0.002	7.521	0.001
	(0.456)	(0.216)	(0.654)	(0.219)	(0.707)
Firm Size	0.103	-74.916	0.042	-75.162	0.044
	(0.003)	(0.037)	(0.332)	(0.040)	(0.297)
Product Competition	0.480	-174.205	-0.029	-186.526	-0.052
	(0.341)	(0.588)	(0.846)	(0.563)	(0.729)
R&D Intensity	-0.000	0.006	0.000	0.006	0.000
	(0.756)	(0.209)	(0.101)	(0.209)	(0.088)
Previous Firm Trials	0.000	0.000	0.000	-0.000	0.000
	(0.505)	(0.997)	(0.474)	(0.998)	(0.426)
Reputation	0.080	-57.740	-0.010	-52.040	-0.007
	(0.131)	(0.004)	(0.326)	(0.000)	(0.474)
Competition (std.)	-0.173	-46.230	0.024	-54.168	0.021
	(0.000)	(0.003)	(0.010)	(0.004)	(0.017)
Reputation x Competition				28.770	0.015
				(0.333)	(0.044)
Constant	-2.879	2,270.727	0.307	2,264.688	0.296
	(0.000)	(0.001)	(0.469)	(0.001)	(0.475)
Observations (Number of Trials)	4,938	3,261	584	3,261	584
Log Pseudo/R-squared	-2071.33	0.401	0.147	0.402	0.149
Number of Firms	396	209	63	209	63
Therapeutic Area FE	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes
Robust SE	Yes	Yes	Yes	Yes	Yes

Notes. Robust *p* values in parentheses. Two-tailed tests.

Appendix C: COX Regression Model for Disclosure Speed.

	(1) Delay Decay
Adverse Events Ratio	0.373
	(0.011)
Phase 3 or 4	0.363
	(0.000)
No of Sponsors	-0.092
	(0.046)
No of Facilities	0.001
	(0.000)
Requested Extension	-1.135
	(0.000)
Number of Subjects	-0.027
	(0.181)
Firm Slack	0.012
	(0.018)
Firm Size	0.104
	(0.000)
Product Competition	-0.165
	(0.078)
R&D Intensity	0.000
	(0.081)
Previous Firm Trials	-0.000
	(0.867)
Reputation	0.108
	(0.000)
Competition (std.)	0.080
	(0.011)
Observations (Number of Trials)	3,249
Log Pseudolikelihood	-22162.27
Time at Risk	2803366
Therapeutic Area FE	Yes
Year FE	Yes
Robust SEs	Yes

Notes. Robust *p* values in parentheses. Two-tailed tests.

Chapter 3: The dynastic duality: Examining the sociocultural capital and nepotism discount of Bollywood dynasts

ABSTRACT

Family producers are the most prevalent organizational form globally. They are driven by the dynastic motive—to perpetuate their wealth and status over generations so that the descendants can enjoy the fruits of their forebears' labor. However, there are conflicting views on whether descendants of these family producers are innovative or constrained by their ancestors' legacies. On the one hand, descendants may inherit status and tacit knowledge from their ancestors and, thus, enjoy the benefits of accessing and bundling coveted resources, allowing them to pursue innovation through tradition. On the other hand, descendants may be constrained by skepticism about nepotism and the unwillingness of the controlling families to professionalize. In this study, we attempt to understand the dynasty's role in family producers to innovate. We show that the dynastic roots of producers in Bollywood (Hindi movie industry) can lead to an inverted U-shaped innovation curve, where sociocultural capital and nepotism discount can both play a role in hindering or enabling innovation. We discuss the implications of our study on innovation capabilities and dynastic producers.

INTRODUCTION

Family producers are the most predominant organizational form globally (Berrone et al., 2020; La Porta, Lopez-De-Silanes, & Shleifer, 1999). They are driven by the "dynastic motive" or "transgenerational intention"—to perpetuate their wealth and status over generations so that the descendants can enjoy the fruits of their forebears' labor (Chua, Chrisman, & Sharma, 1999; Gomez-Mejia, Haynes, Núñez-Nickel, Jacobson, & Moyano-Fuentes, 2007; Huang, Chen, Xu, Lu, & Tam, 2020; Zellweger, Kellermanns, Chrisman, & Chua, 2012). This motive is also vital for endowing them with socio-emotional wealth (SEW)—the "non-financial aspects of the firm that meet the family's affective needs, such as identity, the ability to exercise family influence, and *the perpetuation of the family dynasty*" (Gomez-Mejia et al., 2007: 106 (emphasis added)).

Innovation is a long-term strategic imperative for family producers to survive across generations. However, are descendants innovative? One view predicts that descendants are more innovative than the founding generation because they inherit status and tacit knowledge (cultural capital) from their ancestors and, thus, enjoy the benefits of accessing and bundling coveted resources (Bourdieu, 1984; Bourdieu & Passeron, 1979; Gedajlovic & Carney, 2010; Washington & Zajac, 2005; Woodfield & Husted, 2017). They are able to use their social capital and pursue "innovation through tradition" (De Massis, Frattini, Kotlar, Petruzzelli, & Wright, 2016; Diaz-Moriana, Clinton, Kammerlander, Lumpkin, & Craig, 2020; Erdogan, Rondi, & De Massis, 2020). However, another view predicts that descendants are constrained by their ancestors' legacies (Erdogan et al., 2020). They are discounted because of skepticism by stakeholders in the next generation about nepotism and the unwillingness of the controlling families to professionalize (Anderson, Duru, & Reeb, 2009; Fang, Kotlar, Memili, Chrisman, & De Massis, 2018; Pérez-González, 2006). In addition, "intergenerational deterioration," a combination of negative effort and ability from the descendants towards value creation from tacit knowledge, constrains their ability to innovate (Gedajlovic, Carney, Chrisman, & Kellermanns, 2012; Memili, Fang, & Welsh, 2015).

These conflicting views about the progeny's innovativeness are puzzling. How do we reconcile these opposite perspectives? Perpetuation of the dynasty is a vital driver of family owners' SEW, differentiating their behavior from that of non-family owners (Gomez-Mejia et al., 2007; Huang et al., 2020). However, it is unclear if pursuing dynasty preservation helps or hurts family producers' innovation propensity. In this study, we attempt to understand the (impeding and enabling) role of the dynasty for producers to innovate. We suggest two opposite latent mechanisms for dynastic producers—sociocultural capital and nepotism discount—leading to an inverted U-shaped innovation curve (Haans, Pieters, & He, 2016). The *dynastic roots* of market players—their embeddedness in the social milieu of markets through their ancestral ties—endow the descendants with an inherited standing in social order and a certain familiarity, or 'feel for the game', of the family's trade (Bourdieu, 1984, 1993; Washington & Zajac, 2005). Thus, dynastic roots confer on them deference from market players and endow them with cultural capital—stock of and familiarity with a field's encoded and legitimate cultural resources (Bourdieu, 1984; Cattani, Ferriani, & Lanza, 2017; De Massis et al., 2016; Nonaka, 1994; Washington & Zajac, 2005). However, dynastic roots also engender a discount by audiences who view descendants as a product of nepotism and unworthy of their business (Hall, 1988; Marcus & Hall, 1992). "Scions of long-standing, 'dynastic' family firms face a disjunction between the family firms and modern societal norms" (Stewart, 2003: 385). In addition to the "nepotism stigma" (Padgett, Padgett, & Morris, 2015), later generations' complacency and inability towards knowledge erode their advantage as the environment changes over time (Gedajlovic et al., 2012; Jeong, Kim, & Kim, 2022).

We build our theory by constructing detailed family trees of Bollywood producers. Bollywood is an old industry (the first feature film in 1913) that dynastic producers dominate, and so far, they have successfully foiled Hollywood studios' enthusiastic entries (Bhushan, 2017; Bose, 2006; Jha, 2020). Second, Bollywood is a project-based industry, for which producers have to form "quasifirms" or "latent organizations"—bringing together various key actors as and when a new project demands (Eccles, 1981; Starkey, Barnatt, & Tempest, 2000). These latent organizations help us highlight the social and cultural mechanisms related to dynastic producers.

We theorize that dynastic producers' innovation performance or the propensity to offer atypicality ,i.e., "novel combinations of prior work" (Uzzi, Mukherjee, Stringer, & Jones, 2013: 468), is inverted U shaped by the latent mechanisms of sociocultural capital and nepotism discount. In a highly uncertain project-based industry, dynastic producers can attract coveted resources. Their dynastic roots foster trust in crewmembers against potential vulnerability exploitation (Granovetter, 2005). Further, dynastic producers better absorb cultural tastes due to their early exposure and training in cultural norms, symbols, and language (Bourdieu, 1984; Bourdieu & Passeron, 1979). Their cultural capital (internalized knowledge, competencies, and dispositions) helps them access and recombine the best resources to offer atypicality. However, highly dynastic producers also suffer from nepotism stigma and intergenerational deterioration that we call the "nepotism discount". This nepotism discount impedes their ability to recombine and instills status anxiety, limiting their novel attempts (Bourdieu, 1993; Gomez-Mejia et al., 2007; Jensen, 2008). Our analyses of Bollywood films from the last decade support our hypotheses. The results are robust after we account for endogeneity through a two-stage instrumental variable approach, and we complement our quantitative analysis with personally conducted and publicly available interviews with Bollywood producers and actors.

With our study, we contribute to the literature on family producers, which has highlighted the dynastic intention of family-owned organizations but has rarely explored the role of dynastic roots permeating markets (Gomez-Mejia, Cruz, Berrone, & De Castro, 2011; Gomez-Mejia et al., 2007; Huang et al., 2020). Our results reconcile the contrasting research we have on the innovative behavior by descendants. The literature on heir discount focuses on nepotism in family firms and shortfalls in family talent, citing the lack of formal management training and professional degrees (Pérez-González, 2006; Smith & Amoako-Adu, 1999). The other view highlights the social status and cultural competencies that descendants inherit and learn from a very young age (Bourdieu, 1984; Le Breton–Miller & Miller, 2015; Washington & Zajac, 2005). We argue that the descendants' propensity to innovate is inverted U-shaped because of the opposite latent mechanisms. We also contribute to the cultural and creative industries literature by emphasizing the mechanisms that shape market exchanges (Becker, 1982; Caves, 2000; Jones, Lorenzen, & Sapsed, 2015; Khaire, 2017).

THEORY AND HYPOTHESES

Descendants and Performance

"The dynastic motive," or the goal of passing the business on to the next generation (Casson, 1999), is a defining feature of family owners. It endows families with socio-emotional wealth (SEW: Gomez-Mejia et al., 2011, 2007) that may make them choose trade-offs with non-economic factors in mind. While SEW successfully explains family owners' pursuit of intergenerational continuity and dynastic motives, it falls short on how the dynastic roots of the businesses might serve or hurt subsequent generations.

So far, we have conflicting theoretical views and evidence regarding the descendants' innovation propensity. One view sees descendants as products of 'wasteful' nepotism (Anderson et al., 2009; Pérez-González, 2006; Smith & Amoako-Adu, 1999). The heirs are discounted because they possess inferior human capital and because key stakeholders such as investors and unhappy employees avoid associating with them (Padgett et al., 2015). This view also predicts intergenerational deterioration as the idiosyncratic business knowledge passed on across generations erodes over time (Gedajlovic et al., 2012). On the other hand, we have an opposite theoretical view that predicts a premium for heirs. This view sees heirs inheriting a socially privileged position in the market and being endowed with cultural capital—knowledge of the field's encoded and legitimate culture (Bourdieu, 1984; Le Breton–Miller & Miller, 2015; Washington & Zajac, 2005). Particularly, in the creative and cultural industries, there are good arguments that family producers are more innovative (Bourdieu, 1984; Le Breton–Miller & Miller, 2015). As Bourdieu and Passeron (1979) note, cultural reproduction is a dimension of social reproduction that lets "the inheritors" remain the dominant class.

This study attempts to resolve the contradictory views regarding progeny innovation. We argue that the innovative performance of the dynastic producers is inverted U-shaped, with two underlying latent mechanisms of socio-cultural capital and nepotism discount that we expand on below.

Dynastic Roots, Status, and Cultural Capital

Since the work by Weber (1978) on status (*ständische Lage* or class position), which comes with privileges and exists independently of economic antecedents, organizational sociologists have extensively studied the effects of status and tried to distinguish it from the related but distinct concepts of reputation and legitimacy (Sauder, Lynn, & Podolny, 2012; Stern, Dukerich, & Zajac, 2014). Although the "historical legacy" (irrespective of the merit-based performance) of market players perpetuates their social ranking (Washington & Zajac, 2005), it is often difficult to observe. Thus, status is often inferred through observable social connections or consecration acts, e.g., awards or rankings, and used as a signal for desirable quality by outsiders (Podolny, 1993, 2010). This signal for desirable quality is especially relevant for project-based industries where producers face uncertainty from their potential audiences for each project. For instance, in a project-based industry, key collaborators want to work with high-status players to advance their careers or protect themselves from exploitation (Granovetter, 2005).

In his seminal work, Bourdieu (1984, 1993) further emphasized the role of status in cultural markets. Higher status helps market players reproduce their social hierarchy positions (their *raison d'être*) because their symbolic power is considered legitimate in the beholder's eyes. Further, the cultural capital of producers—their knowledge of the field's encoded and legitimate culture—plays a major role in maintaining their dominant position. In his works, Bourdieu also theorized that symbolic (read, status) and cultural capital are handed down to the next generation either as a legacy or through "pedagogic action." This legacy and knowledge help agents embody *habitus*—their conscious and unconscious dispositions toward their social reality (Bourdieu, 1984).

We build our latent mechanism regarding dynastic roots on the status and cultural capital. Dynastic market players inherit status and acquire cultural capital from their forefathers. Dynastic roots make market players industry insiders by default due to their historical legacy. Thus, dynastic producers inherit the social position and the privileges that come with it, irrespective of merit (Malter, 2014). Consider an example of the difference between status and reputation: "Were votes for George W. Bush in the 2000 presidential elections based more on his reputation (that is, an assessment of his accumulated experiences and actions in public office) or his status (that is, a social rank held by virtue of sharing a name and genetic relationship with a former president) (Washington & Zajac, 2005: 284)?" Thus, in the case of dynastic producers, status deference may be present even before proving their "capability," since the market has little doubt about their "commitment" to their industry for generations (Zuckerman, 2012).

Second, dynastic roots enhance the cultural capital of creative producers (Bourdieu, 1993). The children of industry leaders are habituated into future roles from an early age (Berger & Luckmann, 1966). They are socialized about role-specific knowledge that they internalize as a "feel for the game" or *habitus* (Bourdieu, 1984). This tacit knowledge is unarticulated and tied to intuition (Nonaka, 1994; Nonaka & von Krogh, 2009). It is acquired through close mentorship from elders and is usually idiosyncratic to the specifics of an industry (Bourdieu, 1993). It is "embodied" in habits, skills, and dispositions due to the cultural infusion of descendants into norms, symbols, and language (Bourdieu, 1984). Thus, "habitus is the result of a long process of inculcation, beginning in early childhood, which becomes a 'second sense' or second nature" (Bourdieu, 1993: 5).

In cultural industries, symbolic and cultural capital allow "participants to reproduce their positions—thus influencing the choice of (and return to) different aesthetic strategies" (Cattani et al., 2014: 258). Bourdieu (1993) highlights the role of symbolic and cultural capital in acquiring economic capital in cultural industries. Status and cultural capital (re)produce prestige hierarchies that may shape the industry's evolution and impose handicaps on peripheral market players or "outsiders" (Bourdieu, 1993; Cattani et al., 2017).

Dynastic Roots, Nepotism Stigma, and Intergenerational Deterioration

Nepotism or preferential treatment to (next-generation) family members in the business is considered pejorative since it breaches the link between performance and rewards. "Even in family firms in the so-called Confucian environment of Hong Kong, 'nepotism' is a disparaged phenomenon" (Stewart, 2003: 386). For instance, employees who observe nepotism but do not benefit from it, stigmatize nepots (Padgett et al., 2015). Family owners are conscious of this stigma and try to be "strategic" in their nepotism by practicing it in conjunction with principles of merit (Jeong et al., 2022). "Family firms choose to downplay their family ties due to negative stereotypes associated with family involvement like nepotism and particularism" (Zellweger, Kellermanns, Eddleston, & Memili, 2012: 247).

While stigma regarding nepotism is a *perception* among audiences, especially when they judge through the lens of meritocracy and individualism, it may or may not be a true reflection of the human capital at work, especially in uncertain and weak institutional environments (Firfiray, Cruz, Neacsu, & Gomez-Mejia, 2018). In cultural industries "with a high degree of uncertainty and product market differentiation, tacit knowledge and reputational resources are quite valuable" (Firfiray et al., 2018: 87). Because it is difficult for outsiders to gain access to tacit knowledge, family producers are more likely to engage in the nepotism "gamble" (Firfiray et al., 2018). However, tacit knowledge would only benefit the second generation by being able to learn from the first generation, "which could help bolster opportunities to claim merit based on human capital. When it comes to later-generation family members, the transfer of founder-specific resources weakens substantially. In later generations, the advantages of resource and knowledge transfer are likely to erode, as it is natural for the usefulness of idiosyncratic knowledge and know-how to diminish as environments change over time" (Jeong et al., 2022: 664). In other words, the cultural capital of dynastic producers may become outdated with cultural shifts, leading to intergenerational deterioration (Gedajlovic et al., 2012). Kidwell and colleagues (2018) also argue that the parent-child relationship imprint may manifest into "bad habits" such as culture of entitlement and injustice across generations.

Thus, our second latent mechanism regarding the costs of dynastic roots is built on nepotism stigma among potential stakeholders and the intergenerational decline of managerial and entrepreneurial skills. Our theoretical mechanisms are illustrated in Figure 1.

Insert Figure 1 about here

Bollywood and Dynasties

We test our theoretical model around the dynastic producers in Bollywood, a term used for Bombay's Hindi-language film industry (now Mumbai). Cinema in India encompasses Hindi films and has many production centers throughout India. Even under British rule, India was the third-largest producer of films globally (Kaul, 1998). A Lumiere Brothers representative introduced motion picture technology in India in July 1896, a few months after the first demonstration in Paris in December 1895 (Bose, 2006).

One of the early pioneers of Indian cinema was Dhundiraj Govind Phalke, generally known as Dadasaheb Phalke, or the "Father of Indian Cinema," who produced the first Indian feature film in 1913. Phalke raised money from family and friends (at one point, his wife had to pledge her jewelry for Phalke to secure a loan: Bose, 2006: 49) to acquire equipment from England and set up his facility in Nasik. With his wife, five sons, three daughters, and other relatives, Dadasaheb built a film studio, employing over a hundred employees. The studio produced approximately 44 silent films (with the last film post-synchronized for sound in 1932) (Bose, 2006; Raheja & Kothari, 2004). Phalke's contributions to cinema were recognized with the establishment of the Dadasaheb Phalke Lifetime Achievement Award in 1969 by the Indian government.

India's first talkie was *Alam Ara* (Beauty of the World) in 1931, which, with its seven songs, "established music, song, and dance as staples of Indian cinema" (Ganti, 2004: 11). It also starred Prithviraj Kapoor, who became India's first superstar and established the Kapoor dynasty (sometimes also referred to as the "First Family of Bollywood"). The Kapoor dynasty is still a major force in Bollywood, with the fourth generation active, and has the unique distinction of receiving three Dadasaheb Phalke awards—Prithviraj Kapoor in 1971, Raj Kapoor in 1987, and Sashi Kapoor in 2014 (Barnouw & Krishnaswamy, 1980). We illustrate the Kapoor dynasty tree in Table 1.

Insert Table 1 about here

Although the Indian film industry evolved with Hindi films being produced outside of Bollywood, the dominance of elite families has remained a constant for over a century. "Comprising primarily family businesses, the Hindi film industry has never been vertically or horizontally integrated in the manner of the major Hollywood studios or multinational entertainment conglomerates" (Ganti, 2015: 454). However, the presence of independent producers has not stopped symbolic and cultural capital from asymmetrically moving to industry "heavyweights"—a euphemism for market actors embedded in the industry for generations and having considerable social capital and the power to "validate" careers (Hindustan Times, 2020). This power asymmetry of outsiders' struggles to make it in Bollywood came to the fore again with the unfortunate suicide of an upcoming actor, Sushant Singh Rajput. Although no suicide note was found, the incident led to the umpteenth discussion about the uphill struggle that actors with no family connections to Bollywood face (News18, 2020).

Overall, Bollywood gives us an excellent context for the workings of dynastic producers in a nation obsessed with film personalities (Bose, 2006; Kurzman et al., 2007). As well-known

Bollywood authority Ganti (2004: 55) notes about the role dynasties play, "With most film people marrying other film people and with their children entering the industry, the Bombay film industry appears to be literally reproducing itself." Due to the Bollywood personalities' public lives being under a constant media lens, we can construct their detailed family histories. As another academic researching culture that we interviewed noted: "I am sure someone somewhere has a list of physical deformities of actors, or their shoe sizes, or their various screen names…I have researched across western and Indian culture, and I am yet to see more obsessed (film) fans than Indians."

Bollywood, being a cultural industry, also helps us identify the latent mechanisms of dynastic roots. It is a project-based industry, and for each film, producers have to form a latent organization "that binds together configurations of key actors in ongoing relationships that become active/manifest as and when new projects demand" (Starkey et al., 2000: 299). Thus, for each movie, actors, directors, and other crewmembers have to decide about working with producers. These decisions and their eventual innovative outcome are likely to depend on the social and cultural factors affecting producers. As a thought experiment, if we think of an industry on the other end of the spectrum with a stable environment and an undifferentiated product market, such as commodities or household groceries, the social and cultural mechanisms regarding producers are less likely to play a crucial role.

To gain deeper insights into our context, we also consulted archival sources and conducted interviews to understand the role of dynastic roots in our industry. We started with a well-known dynastic producer and an actor with many critically acclaimed roles. We then used a snowball approach to identify other informants: two producers, two actors, a well-known director, a film editor, an academic, and a film critic.

Descendants Offering Atypicality

Our fundamental argument is that the dynastic producers' propensity to offer atypicality (Elberse, 2007; Goldberg, Hannan, & Kovács, 2016; Hsu, 2006) is inverted U-shaped, i.e., low with both non-and highly dynastic producers. Below, we list latent mechanisms for both directions. First, the inherited social status of dynastic producers acts as a signal of desirable qualities to talent and crew. Due to the project-based work, the producers organize themselves into latent organizations rather than vertically integrating the specialists (Starkey et al., 2000). The vertically integrated studio era in Bollywood was a fleeting decade before Indian independence (Ganti, 2004), significantly shorter than the one in Hollywood that lasted until the mid-1960s (Miller & Shamsie, 1996). The dynastic producers get blocked dates from the best crew and technicians, who sometimes work on multiple projects due to the freelance nature of their work. The dynastic producers sometimes also have long-standing relationships with studios, distributors, and exotic locales where they shoot song and dance sequences—a distinguishing feature of Bollywood films. They are given preferential treatment in various locations, which helps them exploit the *mise en scène* into atypical offerings. For instance, the Swiss government has honored Yash Chopra (from the Chopra-Johar Dynasty) with the title "Ambassador of Interlaken" for his continued filming in Switzerland since the 1970s. Jungfrau Railways also launched a train in his name—an honor shared only with the railway's founder (Bhushan, 2011). Well-known actors are more willing to work with dynastic producers (Chung, Singh, & Lee, 2000; Jensen & Kim, 2015; Podolny, 1993). As Anurag Kashyap (a Bollywood producer and director) reflected, Sushant Singh (the case referred to above) had little choice but to work with a dynastic producer due to his outsider status. "That validation you are seeking, and one can't blame you. It's a choice you have made, but you also have to deal with this. The boy was very talented, but at that time, he chose "Drive" over a film that I was doing because he was dying to work with Dharma (a production company headed by Karan Johar from the Chopra-Johar Dynasty)" (Hindustan Times, 2020).

Second, dynastic producers are better at absorbing and pioneering cultural trends because of their early exposure to cultural norms, symbols, and language (Berger & Luckmann, 1966; Bourdieu, 1984). The "cultural privilege" of being born into a Bollywood family helps dynastic producers to pursue diverse cultural interests, e.g., theater, poetry, and classical music (Bourdieu, 1984). As noted by Shammi Kapoor in an interview (a successful actor and producer from the Kapoor Dynasty and the younger son of Prithviraj Kapoor), "… it took me time, when I joined the theater to start learning—Mr. Amitabh Bachchan's father's books—they were more or less our syllabus in the theater. My father was very adamant about it: *"You must all go through those books of his poetry, and learn the language. Hindi is very important."* I have no pretensions of saying that I was good at it, but I made an effort, and that is very important" (Ganti, 2004: 186–187). Dynastic producers are also likely to get repetitive training and close mentoring from the experienced members of their families (Nonaka & von Krogh, 2009). This apprenticeship is important for developing their own style of "fruitful dilettantism" (Bourdieu & Passeron, 1979). They are also prone to visiting film sets and having informal conversations with film crews about various film intricacies and elements (Bose, 2006). As Pooja Bhatt (a former actress and now a producer-director from the Bhatt dynasty) notes, "…I am basically from a film family where I've always been involved in every aspect of the film, not only coming on to the set and doing my bit and leaving" (Ganti, 2004: 191). As a result, dynastic stars are more likely to cultivate a natural curiosity to experiment (Bourdieu, 1984).

Thus, the sociocultural capital of dynastic producers helps them have long-standing relationships with coveted resource providers, get preferential treatment, and develop deep tacit knowledge of their craft to offer atypicality. However, dynastic roots have their own challenges that we call the "nepotism discount".

First, many potential stakeholders, especially those who have been "outsiders" to the industry and yet found success, want to avoid dynastic producers (Cattani et al., 2014). The non-dynastic players may judge dynastic producers against their meritocratic standards and their struggles to establish themselves in the industry. For instance, Kangana Ranaut, a well-known Bollywood actress, appeared on a popular Indian TV show and called the show host Karan Johar, a film producer/director belonging to one of the most influential families in Bollywood (the Chopra-Johar family), "the flag-bearer of nepotism" who is intolerant of outsiders (India Today, 2017). The dynastic producers are also acutely aware of their "ascribed" or unearned status—"the status that actors inherit based on their association to groups due to nominal characteristics" (Prato, Kypraios, Ertug, & Lee, 2019: 1004). This ascribed status due to an "accident of birth" (Foladare, 1969: 53) differs from the achieved status that producers earn through their accomplishments. As a result, highly dynastic producers are unable to cultivate new connections, which can limit their opportunities to innovate (Chen, Peterson, Phillips, Podolny, & Ridgeway, 2012).

Second, highly dynastic producers will likely suffer from intergenerational deterioration (Gedajlovic et al., 2012). They have inherited competencies and cultural knowledge from their forefathers, who presumably were successful in certain genres of filmmaking. Since the cultural knowledge has worked for their ancestors, the descendants are likely comfortable working with the same conventions and will not be able to appropriate the cultural trends (Bourdieu, 1984). For instance, Sooraj Bharjatya (a producer and director from the Bharjatya Dynasty) is dubbed a "wedding planner" since all his films revolve around elaborate Indian weddings (Jhunjhunwala, 2008). Further, highly dynastic producers will be conventional in their choices, as it helps them claim that they are more "authentic" to their legacy for their audiences (Lehman, O'Connor, Kovács, & Newman, 2019). Thus, the historical legacy of these highly dynastic producers may handcuff them from innovating (Delmestri, Wezel, Goodrick, & Washington, 2020).

In summary, while dynastic producers inherit sociocultural advantages, their dynastic roots also impede their ability to achieve peak innovation. Thus, the dynastic producers at the middle level are the ones who will be the most innovative. Therefore, we expect that *dynastic producers have an inverted U-shaped relationship with atypicality in their offerings.*

DATA AND METHODS

Sample and Data Sources

To test our theoretical model, we assembled a sample of Bollywood movies, their producers and stars' family trees, and their atypicality information from various publicly available sources. We began collecting our data from the annual Bollywood films list curated by *bollywoodhungama.com* (BH). BH is a major Bollywood website that curates films and their various characteristics. We then matched the films from BH to the films individually on *boxofficeindia.com* (BOI)—the go-to website for box office information. Although BH also gives basic box office collection information (albeit with more missing values), there has been criticism that it inflates that number, and BOI is a better source for it with transparent definitions[9] (Khilnani, 2017). We then matched these films to *imdb.com* to obtain genre data, as BH and BOI typically list only one genre.[10] Thus, we can triangulate the data on Bollywood films from various data sources.[11] In total, we are left with a

[9] https://boxofficeindia.com/content.php?pagekey=faq
[10] We confirmed whether IMDB genres overlap with single BOI genres and found that IMDB genres always included the BOI genres, often as the first genre.
[11] Bollywood is generally criticized for poor reporting of data. However, the situation has improved over the last decade as professional market research companies have entered the industry (Dastidar & Elliott, 2020).

panel of 1425 films by 942 (primary)[12] producers from the year 2009 until March 2020, when the cinema halls were ordered to shut down due to COVID-19 lockdowns in most states (BBC, 2020). Next, we collected the family tree and dynastic information about the movies' producers and primary cast members, described in our independent variable section below.

Dependent Variable

Our dependent variable is the *atypicality* of the movies produced (Goldberg et al., 2016; Hsu, Hannan, & Koçak, 2009). Following previous research, we measure it through the number of genres spanned by the movies, considering the level of similarity between those. The rationale behind this approach is that if a film spans multiple genres, the film is unlikely to conform to each genre's typical norms. Further, it takes into account the distances of spanned genres as well. For instance, both romantic-comedy and horror-comedy have a *niche width* (spanned genres) of two (Hsu, 2006). However, a movie combining the horror genre with comedy is more atypical than a movie spanning the romance and comedy genres (proximate genres). Formally:

$$Atypicality = 1 - \left(\frac{1}{1 + \frac{D(x)}{(|l_x| - 1)}} \right), if\ |l_x| > 1$$

where $|l_x|$ denotes the number of genres to which a film belongs, and $D(x)$ denotes the sum of the pairwise cognitive distance between those genres as calculated using an adjusted Jaccard similarity index (Goldberg et al., 2016).

Independent Variable

We constructed the dynastic roots of the producers and actors of our sampled films. We traced all the family members of the actors and producers through their available Wikipedia page sources

[12] A film usually names multiple producers. We take the primary producer to construct our panel. For our producer dynasty variable, we average their dynasty indices as described below.

and mainstream media articles. The first round of construction was done by an external researcher who was a foreigner to the context. Then, in the second round, one co-author familiar with the context worked with the external researcher to finalize the trees.[13] The rules we followed while constructing family trees were: first, we left out family members who were cricket stars, business tycoons, or politicians. Since Bollywood is an attractive industry, many "rich and powerful" enter it for glamour. But counting every rich, consequential, or famous person would dilute our variable since our arguments are about status and cultural capital within the industry. Next, we cut off the dynasties at the spousal level if the two dynasties are related. However, we counted their children together (not stepchildren) in both dynasties. For instance, in the Kapoor Dynasty (Table 1), the fourth generation is active in Bollywood. Kareena Kapoor, a superstar, is married to Saif Ali Khan, from another film dynasty (the Ali-Tagore dynasty).[14] Saif Ali Khan has two kids from his first marriage to actress Amrita Singh—Sara Ali Khan and Ibrahim Ali Khan, who are now active in Bollywood. Thus, we count Sara and Ibrahim in the Ali-Tagore dynasty and Amrita Singh's dynasty but not in the Kapoor dynasty.

Then, for each dynasty, we create two measures: a) *depth*, or the number of generations, and b) *breadth*, or the number of Bollywood people in the dynasties. These two are, of course, highly correlated (0.70). We thus created two alternate measures. First, we add their standardized values with equal weighting, and second, we conduct a PCA (Principal Component Analysis) using the principal factor method with Promax rotation. As expected, the PCA yielded only one factor

[13] We have relevant hyperlinks to articles describing kinship ties wherever applicable.
[14] Saif Ali Khan's father, Mansur Ali Khan, was from an actual ruling dynasty and married to a famous film actress, Sharmila Tagore. Mansur Ali Khan was a "prince" until the privy purses of pre-British times were abolished in the Constitution of India, and the prince titles ceased to exist. He was also a famous cricketer who led the Indian cricket team in the 1960s and 70s. However, as we noted, we do not count him in our family trees.

of the eigenvalue of 1.7, with 85% of the variation explained. Our methodology is similar to Stern and colleagues (2014) study in which a scientist's reputation index is constructed through publication counts and citations. The results were similar for both measures, but for brevity, we report the results obtained by the PCA measure in the main analyses. The results with the alternative measure are reported in Appendix S2.

Finally, we map the dynastic index measure onto the films. If more than one dynastic producer is in a film, we average their dynastic roots measures. Averaging the measure helps us avoid double-counting the dynasty's impact since members of the same family are often named as the main producers.[15] Siblings, parent-children, or couples may have a very similar stock of status or cultural capital. Thus, we have a producers' dynastic index at the film level. We have one dynasty with the fifth consecutive generation active. The maximum number of film personalities in a dynasty is 29 in the Bhatt dynasty (Pooja Bhatt from the dynasty was mentioned earlier in the study).

Control Variables

We controlled for several potential sources of confounds that might affect the relationship between dynastic roots and their ability to offer atypical films. First,we controlled for *superstars* (De Vany & Walls, 1999; Elberse, 2007). We operationalized superstars through the box office success that the stars had in the two-year window before the film's release year. Our underlying logic was that producers would chase today's superstars to sign their films to appear in one to two years—the average time for a firm from conception to release. Thus, the current release should check if the actor was a superstar one or two years prior. We, therefore, take the list of the top ten box office performers (both male and female) from BH in the years t-2 and t-1 and assign them a value of

[15] We repeated our analyses using only the dynasty index of the first-named producer, and our results remained consistent.

one. Next, we controlled for *producer reputation* since their merit-based history might confound the hypothesized relationships of dynastic roots. We operationalized it through previous box office successes that the producers had in the two-year window before the film's release. We counted a box office success if the movie was in the top five box office grossers. More producers in a movie may bring more resources that may affect the hypothesized relationships. Thus, we controlled the *number of producers*. *Top directors* may also drive the relationship between dynastic roots and their ability to offer novelty (Hsu, 2006). Thus, we took the list of the most successful directors from BOI (top 50 in the last two decades) and created a dummy variable (1/0) if the film had a top director.

The movies released during the holiday season (*holiday release*) may also affect the hypothesized relationships (Hsu, 2006). Hence, we create a dummy variable if the movie is released during the weeks of major Indian holidays—New Year's Eve, *Pongal/Makar Sakranti*, Republic Day, *Holi*, Independence Day, *Eid-ul-fitra*, *Dussera*, *Diwali*, and Christmas. The *movie duration* may also confound the relationship between dynastic producers and the movie atypicality. Thus, we control for movie length in our models (in minutes). Finally, we included *year* and *primary genre* dummy variables to control for time fixed-effects such as changing audience tastes and technological changes in the production and distribution of films.

Estimation Methods

We run OLS regression models predicting the atypicality of films. We used "xtreg" command in Stata with robust standard errors clustered around the primary producer. For robustness check, we also run two-stage methods with the instrumental variables predicting the dynastic roots of producers and actors in the first stage. In the second stage, we predict the atypicality of films. We describe the Instrumental variable in the robustness section. Specifically, we use the "xtivreg" command in Stata that incorporates IVs and get similar results. However, we report the OLS regression models below for brevity's sake.

RESULTS

The descriptive statistics and correlations for the sample are reported in Table 2. On average, the films had an atypicality score of 0.53. We computed each model's variance inflation factor (VIF) to determine whether multicollinearity was biasing our estimates. We discovered that the greatest VIF statistics were 2.08 and the average VIF value was 1.33 (Hair, Anderson, Tatham, & Black, 1995), indicating that multicollinearity was not a problem.

Insert Tables 2-3 & Figures 2-3 about here.

Table 3 reports the results of the effect of the producer dynasty on atypicality. Model 1 is with all the control variables. In the interests of parsimony, year and primary genre dummy variables are not shown. In Model 2, the producer dynasty index produced a positive and significant coefficient as a determinant of atypicality ($b = 0.024$, $p < 0.01$) and the coefficient for the squared term of producer dynasty (*producer dynasty^2*) was negative and significant ($b = -0.004$, $p < 0.05$), suggesting an inverted-U relationship. Second, we evaluated the slopes at the lowest and highest values of the dynasty index and found that the slope of it is positive and significant ($b = 0.041$, $p <. 01$) at the lowest level and negative and moderately significant ($b = -0.035$, $p <. 1$) at the highest level. Finally, we calculated the turning point of the curve, the point at which dynasty yields the highest atypicality (3.08), which was well within the range of the dynasty index in our data. We thus conclude that having dynastic roots enhances atypicality until a point after which the effect turns negative, supporting our main prediction (Haans et al., 2016). We illustrate the average marginal effect in Figure 2.

Additional Analysis

In Model 4, we interacted the producer dynasty index and its square term with actor dynasty index. To illustrate the relationship, we estimated atypicality at different levels of producer dynasty (within the sample's 1^{st} and 99^{th} percentile) and actor dynasty (median and one standard deviation below/above the mean). The plot in Figure 3 suggests that dynastic actors flatten the curve. Model 4 also shows that the interaction term between actor dynasty and producer dynasty produced a negative and significant coefficient ($b = -0.008$, $p <. 01$), while the interaction term between actor dynasty and producer dynasty^2 produced a positive and moderately significant coefficient ($b = 0.001$, $p <. 1$). We further tested for a shift in the turning point for the curve but found the coefficient of the suggested equation by Haans et al (2016) not significantly different from zero at meaningful values of actor dynasty index. Thus, evidence points to a flattening of inverted U-shape, suggesting that dynastic actors increase the nepotism discount for dynastic producers without benefitting their sociocultural capital.

Robustness Tests

In our data, we may have omitted variables that may drive the hypothesized relationships. Hence, we constructed a variable to instrument our main independent variable of dynastic roots. We take the ratio of male children to total children of the first generation. Note that we take all the male and female children (not just those who entered the film industry). We assume that more male descendants are likely to enter the film industry than female descendants. This is not a very farfetched assumption. Bollywood has traditionally been a bastion of male actors and producers. When Dadasahab Phalke tried to cast females for India's first feature film, women were unwilling to work due to the stigma attached to films. "The public nature of the filmic image appeared to violate the dominant norms of feminine modesty…. Even prostitutes were unwilling to act in films since that would appear as a public disclosure of their occupation, so in Phalke's first film, *Raja Harischandra*, the role of Queen Taramati was played by a young man" (Ganti, 2004: 12).

Although times have changed, Indian families still discourage their women from entering the film industry. The Kapoor dynasty (from Table 1) was known for not letting their daughters enter films, and if a son married an actress, the wife would stop acting. This practice carried on until the current generation of Kareena Kapoor. Her elder sister, Karishma Kapoor, was the first daughter of the Kapoor clan to enter the film industry, and her family ties were estranged for a time. Thus, we use the randomness of children's sex ratio as an instrument for the dynasty. The exclusion assumption of the instrument variable is that any effect on the atypicality is through the supply shock (or randomized effect) it generates on the dynasty.

One argument against our instrument is that what if the sex of the children was not random? In India, female infanticide (although illegal) has been a problem, but in the poorer sections of society (Bhatnagar, Dube, & Dube, 2012). Further, prenatal sex discernment has been illegal through the prohibition of the self-selection act since 1994 (Jaising, 2004). Finally, the Bollywood people live under the public gaze and constant media lens, and thus, it would be difficult for them to hide their female kids and easy for us to count them. Researchers have used similar instruments. For instance, Bennedsen and colleagues (2007) use the gender of the firstborn child (a random event) to instrument CEO succession decisions by family firms. Their argument is similar: boy-first families are more likely to appoint their (male) CEOs rather than look outside the family for candidates. Overall, we think the sex ratio is a valid instrument for the dynasty and our empirical tests agree. The results from two-stage regressions were qualitatively similar to our OLS results.

DISCUSSION

Our results in Bollywood show that the dynastic roots endow producers with sociocultural capital and nepotism discount that both enable and constrain their innovation efforts. These results have several theoretical and practical implications.

Contributions

Our contribution is to the family business literature that has underlined the dynastic motive of business families but has yet to resolve whether dynastic roots are pain or gain for them (Casson, 1999; Gomez-Mejia et al., 2011). The conflicting results about the sociocultural capital and nepotism discount of heirs are puzzling. The studies on heir discount show that external stakeholders worry about the lack of professional education and nepotism in the next generation of leadership (Pérez-González, 2006). Recently, when Bernard Arnault, chief at LMVH, appointed his eldest daughter as the head of Dior, the critics were quick to call her "nepobaby" (Abboud, 2023). Some of these stigmatizations of nepotism may be unjustified and a knee-jerk reaction on the part of various stakeholders. Delphine Arnault has been a key executive at LMVH since 2013 and has acquired the training and experience to lead a major luxury group. The nepotism stigma assumes that knowledge is only transferred through formal institutions. Family firms are vital informal institutions that transfer knowledge and competencies from an early age that the next generations can leverage as resources (De Massis et al., 2016). Bourdieu (1984), in his work on education, highlighted the pedagogy that the families employed in transferring their cultural capital.

Not only cultural capital but also social capital is transferred through the generations. The historical legacy of the market players endows them with higher status. Typically, inherited historical legacies are difficult to quantify and must be inferred through market participants' embeddedness or social rankings (Sauder et al., 2012; Washington & Zajac, 2005). By constructing the family trees, we could test the assumption of historical legacy that the literature on status makes. This high status of dynastic market players signals desirable qualities to relevant audiences.

The "status as a signal" of desirable quality (Podolny, 1993) is critical in industries that operate with freelancers or on a project basis, such as construction, consulting, design, high-tech, and the entertainment industry. Over the last few decades, more and more industries have shifted to short-term, contractual, project-based organizations (Manning, 2017). In such industries, considerable uncertainty about the focal market actor among external stakeholders persists. By their existence, dynastic producers have shown commitment to an industry and are likely to be major players in its future, are in an advantageous position. They can use their symbolic power to attract the best resources and erect further barriers to entry for competitors.

Further, in cultural and creative markets, there is considerable uncertainty about the quality of the final product, which Caves (2000) elegantly called "nobody knows" (or symmetrical ignorance). These are not standardized products that producers can control through formal quality checks. The freelancers in the entertainment industry do their value-add without knowing how the project will further evolve. Again, the dynastic producers have an advantage due to their cultural capital. With the big picture in mind, they can effectively orchestrate the best resources (Duran, Kammerlander, van Essen, & Zellweger, 2016). However, their cultural capital can also get outdated, especially in a fast-changing cultural industry. In addition, the pressure to maintain their legacy can prevent them from being innovative (Diaz-Moriana et al., 2020). Overall, we are able to reconcile the opposite theoretical mechanisms in favor of and against dynastic producers' quest for innovation.

We also discovered that collaborations between dynastic producers and actors exacerbated the nepotism discount while not benefiting sociocultural capital. The dynastic actors feel pressured by audience expectations to regurgitate their ancestor's work. For instance, the son of Bollywood superstar Amitabh Bachchan, Abhishek (of the Bachchan Dynasty), was criticized for copying his father. "The audience wants to see Abhishek. Not a poor copy of Amitabh" (Chopra, 2000).

Limitations, Future Directions, and Conclusion

Since this work is the first attempt to reconcile the innovative behavior of dynastic producers, it has several limitations that future research can address. First, we were limited by the data availability of films from the last decade in our context. In other family-dominated industries for generations, such as the circus industry, it may be possible to understand how novelty varies over the generations. Second, although we tried to reach causality with the IV approach, our setting is not a controlled lab. Future research can explore the mechanisms in controlled experiments. Finally, qualitative studies may also look at the micro-processes that dynastic players employ in their informal dealings or with lower-status groups.

Despite the limitations, we could push our understanding of dynastic market players. The duality of dynastic players helps and hurts their innovation projects simultaneously. The sociocultural capital and nepotism discount work in opposite directions, which dynastic producers need to be aware of and find an optimum point. They need to update their cultural capital and manage their perceptions proactively with their potential stakeholders to innovate.

REFERENCES

Abboud, L. 2023, January 11. Bernard Arnault promotes daughter Delphine in LVMH reshuffle. ***Financial Times***.

Anderson, R. C., Duru, A., & Reeb, D. M. 2009. Founders, heirs, and corporate opacity in the United States. ***Journal of Financial Economics***, 92(2): 205–222.

Barnouw, E., & Krishnaswamy, S. 1980. ***Indian film***. New York: Oxford University Press.

BBC. 2020, October 16. India Covid-19: Bollywood faces biggest box office test as cinemas open. ***BBC News***. https://www.bbc.com/news/world-asia-india-54565884.

Becker, H. S. 1982. ***Art Worlds***. Berkeley, CA: University of California Press.

Bennedsen, M., Nielsen, K. M., Perez-Gonzalez, F., & Wolfenzon, D. 2007. Inside the Family Firm: The Role of Families in Succession Decisions and Performance*. ***The Quarterly Journal of Economics***, 122(2): 647–691.

Berger, P. L., & Luckmann, T. 1966. ***The social construction of reality. A treatise in the sociology of knowledge.*** New York: Anchor Books.

Berrone, P., Duran, P., Gómez-Mejía, L., Heugens, P. P. M. A. R., Kostova, T., et al. 2020. Impact of informal institutions on the prevalence, strategy, and performance of family firms: A meta-analysis. ***Journal of International Business Studies***. https://doi.org/10.1057/s41267-020-00362-6.

Bhatnagar, R. D., Dube, Renu, & Dube, Reena. 2012. ***Female infanticide in India: A feminist cultural history***. New York: SUNY Press.

Bhushan, N. 2011, April 12. Indian Filmmaker Yash Chopra Receives Swiss Honors. ***The Hollywood Reporter***. https://www.hollywoodreporter.com/news/general-news/indian-filmmaker-yash-chopra-receives-177399/.

Bhushan, N. 2017, January 10. Disney India to Cut Jobs Amid Exit From Bollywood Productions. ***The Hollywood Reporter***. https://www.hollywoodreporter.com/news/general-news/disney-india-cut-jobs-exit-bollywood-productions-963059/.

Bose, M. 2006. ***Bollywood: A History***. UK: Tempus.

Bourdieu, P. 1984. ***Distinction: A Social Critique of the Judgement of Taste***. Cambridge, Massachusetts: Harvard University Press.

Bourdieu, P. 1993. ***The field of cultural production: Essays on art and literature***. Columbia University Press.

Bourdieu, P., & Passeron, J. C. 1979. ***The inheritors: French students and their relation to culture***. Chicago: Univ of Chicago Press.

Casson, M. 1999. The economics of the family firm. ***Scandinavian Economic History Review***, 47(1): 10–23.

Cattani, G., Ferriani, S., & Allison, P. D. 2014. Insiders, Outsiders, and the Struggle for Consecration in Cultural Fields. ***American Sociological Review***, 79(2): 258–281.

Cattani, G., Ferriani, S., & Lanza, A. 2017. Deconstructing the Outsider Puzzle: The Legitimation Journey of Novelty. ***Organization Science***, 28(6): 965–992.

Caves, R. E. 2000. ***Creative industries: Contracts between art and commerce***. Cambridge, Massachusetts: Harvard University Press.

Chen, Y.-R., Peterson, R. S., Phillips, D. J., Podolny, J. M., & Ridgeway, C. L. 2012. Introduction to the Special Issue: Bringing Status to the Table—Attaining, Maintaining, and Experiencing Status in Organizations and Markets. ***Organization Science***, 23(2): 299–307.

Chopra, A. 2000, November 9. Bollywood: Though box-office glory eludes Abhishek Bachchan, nobody's writing him off. ***India Today***. https://www.indiatoday.in/magazine/society-the-arts/films/story/20000911-bollywood-though-box-office-glory-eludes-abhishek-bachchan-nobodys-writing-him-off-778010-2000-09-11.

Chua, J. H., Chrisman, J. J., & Sharma, P. 1999. Defining the Family Business by Behavior. ***Entrepreneurship Theory and Practice***, 23(4): 19–39.

Chung, S. (Andy), Singh, H., & Lee, K. 2000. Complementarity, status similarity and social capital as drivers of alliance formation. ***Strategic Management Journal***, 21(1): 1–22.

Dastidar, S. G., & Elliott, C. 2020. The Indian film industry in a changing international market. ***Journal of Cultural Economics***, 44(1): 97–116.

De Massis, A., Frattini, F., Kotlar, J., Petruzzelli, A. M., & Wright, M. 2016. Innovation Through Tradition: Lessons From Innovative Family Businesses and Directions for Future Research. ***Academy of Management Perspectives***, 30(1): 93–116.

De Vany, A., & Walls, W. D. 1999. Uncertainty in the movie industry: Does star power reduce the terror of the box office? ***Journal of Cultural Economics***, 23(4): 285–318.

Delmestri, G., Wezel, F. C., Goodrick, E., & Washington, M. 2020. The Hidden Paths of Category Research: Climbing new heights and slippery slopes. ***Organization Studies***, 41(7): 909–920.

Diaz-Moriana, V., Clinton, E., Kammerlander, N., Lumpkin, G. T., & Craig, J. B. 2020. Innovation Motives in Family Firms: A Transgenerational View. ***Entrepreneurship Theory and Practice***, 44(2): 256–287.

Duran, P., Kammerlander, N., van Essen, M., & Zellweger, T. 2016. Doing More with Less: Innovation Input and Output in Family Firms. ***Academy of Management Journal***, 59(4): 1224–1264.

Durand, R., Granqvist, N., & Tyllström, A. 2017. ***From categories to categorization: Studies in sociology, organizations and strategy at the crossroads***. Bingley: UK: Emerald Group Publishing.

Eccles, R. G. 1981. The quasifirm in the construction industry. ***Journal of Economic Behavior & Organization***, 2(4): 335–357.

Elberse, A. 2007. The Power of Stars: Do Star Actors Drive the Success of Movies? ***Journal of Marketing***, 71(4): 102–120.

Erdogan, I., Rondi, E., & De Massis, A. 2020. Managing the Tradition and Innovation Paradox in Family Firms: A Family Imprinting Perspective. ***Entrepreneurship Theory and Practice***, 44(1): 20–54.

Fang, H., Kotlar, J., Memili, E., Chrisman, J. J., & De Massis, A. 2018. The pursuit of international opportunities in family firms: Generational differences and the role of knowledge-based resources. ***Global Strategy Journal***, 8(1): 136–157.

Fini, R., Jourdan, J., & Perkmann, M. 2017. Social Valuation Across Multiple Audiences: The Interplay of Ability and Identity Judgments. ***Academy of Management Journal***, 61(6): 2230–2264.

Firfiray, S., Cruz, C., Neacsu, I., & Gomez-Mejia, L. R. 2018. Is nepotism so bad for family firms? A socioemotional wealth approach. ***Human Resource Management Review***, 28(1): 83–97.

Foladare, I. S. 1969. A Clarification of "Ascribed Status" and "Achieved Status." ***The Sociological Quarterly***, 10(1): 53–61.

Ganti, T. 2004. ***Bollywood: A Guidebook to Popular Hindi Cinema***. London: Routledge.

Ganti, T. 2015. Fuzzy Numbers: The Productive Nature of Ambiguity in the Hindi Film Industry. ***Comparative Studies of South Asia, Africa and the Middle East***, 35(3): 451–465.

Gedajlovic, E., & Carney, M. 2010. Markets, Hierarchies, and Families: Toward a Transaction Cost Theory of the Family Firm. ***Entrepreneurship Theory and Practice***, 34(6): 1145–1172.

Gedajlovic, E., Carney, M., Chrisman, J. J., & Kellermanns, F. W. 2012. The Adolescence of Family Firm Research: Taking Stock and Planning for the Future. ***Journal of Management***, 38(4): 1010–1037.

Goldberg, A., Hannan, M. T., & Kovács, B. 2016. What Does It Mean to Span Cultural Boundaries? Variety and Atypicality in Cultural Consumption. ***American Sociological Review***, 81(2): 215–241.

Gomez-Mejia, L. R., Cruz, C., Berrone, P., & De Castro, J. 2011. The Bind that Ties: Socioemotional Wealth Preservation in Family Firms. ***Academy of Management Annals***, 5(1): 653–707.

Gomez-Mejia, L. R., Haynes, K. T., Núñez-Nickel, M., Jacobson, K. J. L., & Moyano-Fuentes, J. 2007. Socioemotional Wealth and Business Risks in Family-controlled Firms: Evidence from Spanish Olive Oil Mills. ***Administrative Science Quarterly***, 52(1): 106–137.

Granovetter, M. 2005. The Impact of Social Structure on Economic Outcomes. ***Journal of Economic Perspectives***, 19(1): 33–50.

Haans, R. F. J., Pieters, C., & He, Z.-L. 2016. Thinking about U: Theorizing and testing U- and inverted U-shaped relationships in strategy research. ***Strategic Management Journal***, 37(7): 1177–1195.

Hair, J. F., Anderson, R. E., Tatham, R. L., & Black, W. C. 1995. ***Multivariate data analysis***. NY, US: Macmillan.

Hall, P. D. 1988. A Historical Overview of Family Firms in the United States. ***Family Business Review***, 1(1): 51–68.

Hindustan Times. 2020, July 24. ***Anurag Kashyap says Sushant Singh Rajput chose Drive over his film, claims actor 'wanted validation from Yash Raj Films.'*** https://bit.ly/3EvDWVh.

Hsu, G. 2006. Jacks of All Trades and Masters of None: Audiences' Reactions to Spanning Genres in Feature Film Production. ***Administrative Science Quarterly***, 51(3): 420–450.

Hsu, G., Hannan, M. T., & Koçak, Ö. 2009. Multiple Category Memberships in Markets: An Integrative Theory and Two Empirical Tests. ***American Sociological Review***, 74(1): 150–169.

Huang, X., Chen, L., Xu, E., Lu, F., & Tam, K.-C. 2020. Shadow of the Prince: Parent-incumbents' Coercive Control over Child-successors in Family Organizations. ***Administrative Science Quarterly***, 65(3): 710–750.

India Today. 2017, February 20. Koffee With Karan 5: Kangana takes a dig at Karan, calls him a flag-bearer of nepotism. ***India Today***. https://www.indiatoday.in/movies/celebrities/story/koffee-with-karan-5-kangana-ranaut-bollywood-nepotism-kjo-961580-2017-02-20.

Jaising, I. 2004. ***Pre-conception & Pre-natal Diagnostic Techniques Act: A Users Guide to the Law***. New Delhi: Universal Law Publishing.

Jensen, M. 2008. The Use of Relational Discrimination to Manage Market Entry: When Do Social Status and Structural Holes Work Against You? ***Academy of Management Journal***, 51(4): 723–743.

Jensen, M., & Kim, H. 2015. The Real Oscar Curse: The Negative Consequences of Positive Status Shifts. ***Organization Science***, 26(1): 1–21.

Jeong, S.-H., Kim, Heechun, & Kim, Hicheon. 2022. Strategic Nepotism in Family Director Appointments: Evidence from Family Business Groups in South Korea. ***Academy of Management Journal***, 65(2): 656–682.

Jha, L. 2020, November 3. Hollywood studios relook at India strategies. ***Mint***. https://www.livemint.com/news/india/hollywood-studios-relook-at-india-strategies-11604398386913.html.

Jhunjhunwala, U. 2008, September 30. I would like to leave behind good cinema: Barjatya. ***Hindustan Times***. https://www.hindustantimes.com/india/i-would-like-to-leave-behind-good-cinema-barjatya/story-YJ4MSp7JaD4PuncOpWW7MK.html.

Jones, C., Lorenzen, M., & Sapsed, J. 2015. Creative Industries: A Typology of Change. In C. Jones, M. Lorenzen, & J. Sapsed (Eds.), ***The Oxford Handbook of Creative Industries***. Oxford, UK: Oxford University Press.

Kaul, G. 1998. ***Cinema and the Indian freedom struggle: Covering the subcontinent***. New Delhi: Sterling Publishers.

Khaire, M. 2017. ***Culture and Commerce: The Value of Entrepreneurship in Creative Industries***. Stanford: Stanford University Press.

Khilnani, R. 2017. Box-Office Collection: Who can you trust? | Hindi Movie News—Times of India. ***The Times of India***. https://bit.ly/3Jp7o34.

Kidwell, R. E., Eddleston, K. A., & Kellermanns, F. W. 2018. Learning bad habits across generations: How negative imprints affect human resource management in the family firm. ***Human Resource Management Review***, 28(1): 5–17.

Kurzman, C., Anderson, C., Key, C., Lee, Y. O., Moloney, M., et al. 2007. Celebrity Status*. ***Sociological Theory***, 25(4): 347–367.

La Porta, R., Lopez-De-Silanes, F., & Shleifer, A. 1999. Corporate Ownership Around the World. ***The Journal of Finance***, 54(2): 471–517.

Lamont, M. 2012. Toward a Comparative Sociology of Valuation and Evaluation. ***Annual Review of Sociology***, 38(1): 201–221.

Le Breton–Miller, I., & Miller, D. 2015. The Arts and Family Business: Linking Family Business Resources and Performance to Industry Characteristics. ***Entrepreneurship Theory and Practice***, 39(6): 1349–1370.

Lehman, D. W., O'Connor, K., Kovács, B., & Newman, G. E. 2019. Authenticity. ***Academy of Management Annals***, 13(1): 1–42.

Malter, D. 2014. On the Causality and Cause of Returns to Organizational Status: Evidence from the Grands Crus Classés of the Médoc. ***Administrative Science Quarterly***, 59(2): 271–300.

Manning, S. 2017. The rise of project network organizations: Building core teams and flexible partner pools for interorganizational projects. ***Research Policy***, 46(8): 1399–1415.

Marcus, G. E., & Hall, P. D. 1992. ***Lives in trust: The fortunes of dynastic families in late twentieth-century America***. Boulder: Westview Press.

Memili, E., Fang, H. C., & Welsh, D. H. B. 2015. Value creation and value appropriation in innovation process in publicly-traded family firms. ***Management Decision***, 53(9): 1921–1952.

Miller, D., & Shamsie, J. 1996. The Resource-Based View of the Firm in Two Environments: The Hollywood Film Studios From 1936 to 1965. ***Academy of Management Journal***, 39(3): 519–543.

News18. 2020, June 17. ***Sushant Singh Rajput's Death Sparks Outsider vs Insider Debate in Bollywood***. https://www.news18.com/news/movies/sushant-singh-rajputs-death-sparks-outsider-vs-insider-debate-in-bollywood-2673353.html.

Nonaka, I. 1994. A Dynamic Theory of Organizational Knowledge Creation. ***Organization Science***, 5(1): 14–37.

Nonaka, I., & von Krogh, G. 2009. Perspective—Tacit Knowledge and Knowledge Conversion: Controversy and Advancement in Organizational Knowledge Creation Theory. ***Organization Science***, 20(3): 635–652.

Padgett, M. Y., Padgett, R. J., & Morris, K. A. 2015. Perceptions of Nepotism Beneficiaries: The Hidden Price of Using a Family Connection to Obtain a Job. ***Journal of Business and Psychology***, 30(2): 283–298.

Pérez-González, F. 2006. Inherited Control and Firm Performance. ***American Economic Review***, 96(5): 1559–1588.

Podolny, J. M. 1993. A Status-Based Model of Market Competition. ***American Journal of Sociology***, 98(4): 829–872.

Podolny, J. M. 2010. ***Status Signals: A Sociological Study of Market Competition***. Princeton, NJ: Princeton University Press.

Pontikes, E. G. 2012. Two Sides of the Same Coin. ***Administrative Science Quarterly***, 57(1): 81–118.

Raheja, D., & Kothari, J. 2004. ***Indian Cinema, the Bollywood Saga***. New Delhi: Roli Books.

Sauder, M., Lynn, F., & Podolny, J. M. 2012. Status: Insights from Organizational Sociology. ***Annual Review of Sociology***, 38(1): 267–283.

Scott, W. R. 2005. Great minds in management: The process of theory development. ***Institutional theory: Contributing to a theoretical research program***, vol. 37: 460–484. Oxford, UK: Oxford University Press.

Sippy, P. 2021, December 22. The Maasai TikTokers wowing Bollywood fans. ***BBC News***. https://www.bbc.com/news/world-africa-59690317.

Smith, B. F., & Amoako-Adu, B. 1999. Management succession and financial performance of family controlled firms. ***Journal of Corporate Finance***, 5(4): 341–368.

Starkey, K., Barnatt, C., & Tempest, S. 2000. Beyond Networks and Hierarchies: Latent Organizations in the U.K. Television Industry. ***Organization Science***, 11(3): 299–305.

Stern, I., Dukerich, J. M., & Zajac, E. 2014. Unmixed signals: How reputation and status affect alliance formation. ***Strategic Management Journal***, 35(4): 512–531.

Stewart, A. 2003. Help One Another, Use One Another: Toward an Anthropology of Family Business. ***Entrepreneurship Theory and Practice***, 27(4): 383–396.

Taeuscher, K., Bouncken, R. B., & Pesch, R. 2020. Gaining Legitimacy by Being Different: Optimal Distinctiveness in Crowdfunding Platforms. ***Academy of Management Journal***. https://doi.org/10.5465/amj.2018.0620.

Washington, M., & Zajac, E. J. 2005. Status Evolution and Competition: Theory and Evidence. ***Academy of Management Journal***, 48(2): 282–296.

Weber, M. 1978. ***Economy and society: An outline of interpretive sociology***. Berkeley, CA: Univ of California Press.

Woodfield, P., & Husted, K. 2017. Intergenerational knowledge sharing in family firms: Case-based evidence from the New Zealand wine industry. ***Journal of Family Business Strategy***, 8(1): 57–69.

Zellweger, T. M., Kellermanns, F. W., Chrisman, J. J., & Chua, J. H. 2012. Family Control and Family Firm Valuation by Family CEOs: The Importance of Intentions for Transgenerational Control. ***Organization Science***, 23(3): 851–868.

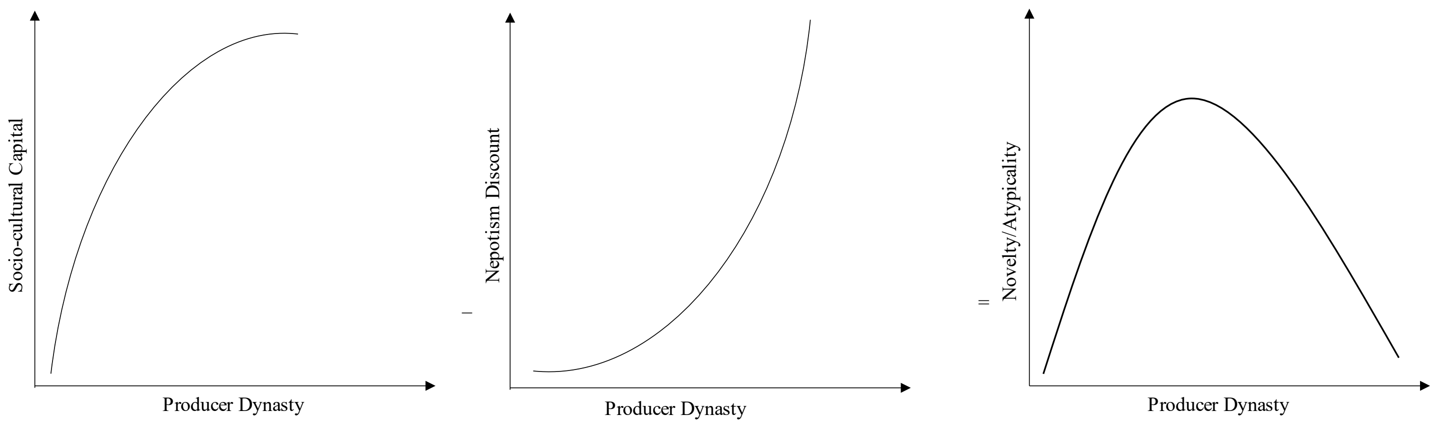
Zellweger, T. M., Kellermanns, F. W., Eddleston, K. A., & Memili, E. 2012. Building a family firm image: How family firms capitalize on their family ties. ***Journal of Family Business Strategy***, 3(4): 239–250.

Zuckerman, E. W. 2012. Construction, Concentration, and (Dis)Continuities in Social Valuations. ***Annual Review of Sociology***, 38(1): 223–245.

Zuckerman, E. W. 2017. The Categorical Imperative Revisited: Implications of Categorization as a Theoretical Tool. ***From Categories to Categorization: Studies in Sociology, Organizations and Strategy at the Crossroads (Research in the Sociology of Organizations)***, vol. 51: 31–68. Emerald.

TABLES AND FIGURES

FIGURE 1



Note: We theorize that the socio-cultural capital mechanism takes a concave shape: the dynastic benefits (status and idiosyncrat ic knowledge) increase at a decreasing rate and eventually level off, while nepotism discount mechanism (stigma and intergenerational deterioration) takes a convex shape : the dynastic costs increase at an increasing rate.

Table 1 – Kapoor Dynasty of Bollywood

Generation 1	Generation 2	Generation 3	Generation 4	Career	Relation
Prithviraj Kapoor				Actor	Son of Dewan Basheshwarnath Kapoor and father of Raj Kapoor
	Raj Kapoor			Actor, producer, director	Son of Prithviraj Kapoor and married to Krishna Kapoor (Elder Son)
		Randhir Kapoor		Actor, producer, director	Son of Raj Kapoor and married to Babita
		Babita		Actress	Married to Randhir Kapoor, daughter of character artist Hari Shivdasani
			Karisma Kapoor	Actress	Daughter of Randhir Kapoor and Babita
			Kareena Kapoor	Actress	Daughter of Randhir Kapoor and Babita
			Saif Ali Khan	Actor	Married to Kareena Kapoor, from Pataudi family
		Rishi Kapoor		Actor, producer, director	Son of Raj Kapoor and married to Neetu Singh
		Neetu Singh		Actress	Married to Rishi Kapoor
			Ranbir Kapoor	Actor	Son of Rishi Kapoor and Neetu Singh
		Rajiv Kapoor		Actor, producer, director	Son of Raj Kapoor
	Shammi Kapoor			Actor, producer, director	Son of Prithviraj Kapoor
	Geeta Bali			Actress	1st wife of Shammi Kapoor
		Aditya Raj Kapoor		Actor, producer	Son of Shammi Kapoor and Geeta Bali
		Kanchan Desai		Actress, Producer	Daughter of Shammi Kapoor and Geeta Bali
	Shashi Kapoor			Actor, producer, director	Son of Prithviraj Kapoor
	Jennifer Kendal			Actor	Wife of Shashi Kapoor
		Kunal Kapoor		Actor, Ad-Film Director	Son of Shashi Kapoor and Jennifer Kendal
		Karan Kapoor		Actor, model, photographer	Son of Shashi Kapoor and Jennifer Kendal
		Sanjana Kapoor		Actress	Daughter of Shashi Kapoor and Jennifer Kendal
		Aditya Bhattacharya		Director, Screenwriter	1st husband of Sanjana Kapoor
		Valmik Thapar		Writer, producer	2nd husband of Sanjana Kapoor
			Armaan Jain	Actor	Son of Reema and Minoj
			Aadar Jain	Actor	Son of Reema and Minoj
Trilok Kapoor				Actor	Brother of Pritviraj
	Vijay Kapoor			Actor, director	Son of Trilok

Table 2 – Summary Statistics and Correlations

Variables	Mean	Std. Dev.	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
(1) Atypicality	0,53	0,4	1									
(2) Producer Dynasty	-0,86	1,98	0,29	1								
(3) Producer Dynasty^2	4,65	6,05	0,04	0,45	1							
(4) Actor Dynasty	-0,19	1,7	0,25	0,39	0,2	1						
(5) Superstar	0,22	0,41	0,32	0,37	0,01	0,3	1					
(6) Producer Reputation	0,15	0,56	0,17	0,41	0,12	0,17	0,26	1				
(7) No. of Producers	1,74	1,22	0,08	0,27	-0,13	0,14	0,11	0,23	1			
(8) Top Director	0,1	0,3	0,22	0,28	0,06	0,22	0,34	0,09	0,04	1		
(9) Holiday Release	0,3	0,46	0,01	0,02	-0,01	0,03	0,03	0	0,03	-0,03	1	
(10) Runtime	122,24	23,97	0,21	0,25	0	0,2	0,31	0,17	0,09	0,26	0,01	1

Note: Correlation coefficients with an absolute value larger than 0.08 are significant at the .01 level.

Table 3 – Multivariate Regressions

Variables	(1)	(2)	(3)
Producer Dynasty		0.024***	0.026***
		(0.008)	(0.008)
Producer Dynasty^2		-0.004**	-0.003
		(0.002)	(0.002)
Actor Dynasty	0.021***	0.019***	0.008
	(0.006)	(0.006)	(0.009)
Producer Dynasty			-0.008***
x Actor Dynasty			(0.003)
Producer Dynasty^2			0.001*
x Actor Dynasty			(0.001)
Superstar	0.119***	0.106***	0.109***
	(0.025)	(0.026)	(0.027)
Producer Reputation	0.023	0.012	0.011
	(0.015)	(0.014)	(0.014)
No of producers	0.014	0.006	0.005
	(0.009)	(0.009)	(0.009)
Top director	0.068***	0.059***	0.065***
	(0.022)	(0.021)	(0.021)
Holiday release	0.003	0.000	0.002
	(0.021)	(0.020)	(0.020)
Runtime	0.001***	0.001***	0.001***
	(0.000)	(0.000)	(0.000)
Constant	0.334***	0.430***	0.426***
	(0.107)	(0.111)	(0.111)
Observations	1,425	1,425	1,425
Number of Primary Producers	942	942	942
Year FE	Yes	Yes	Yes
Genre FE	Yes	Yes	Yes

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

FIGURE 2

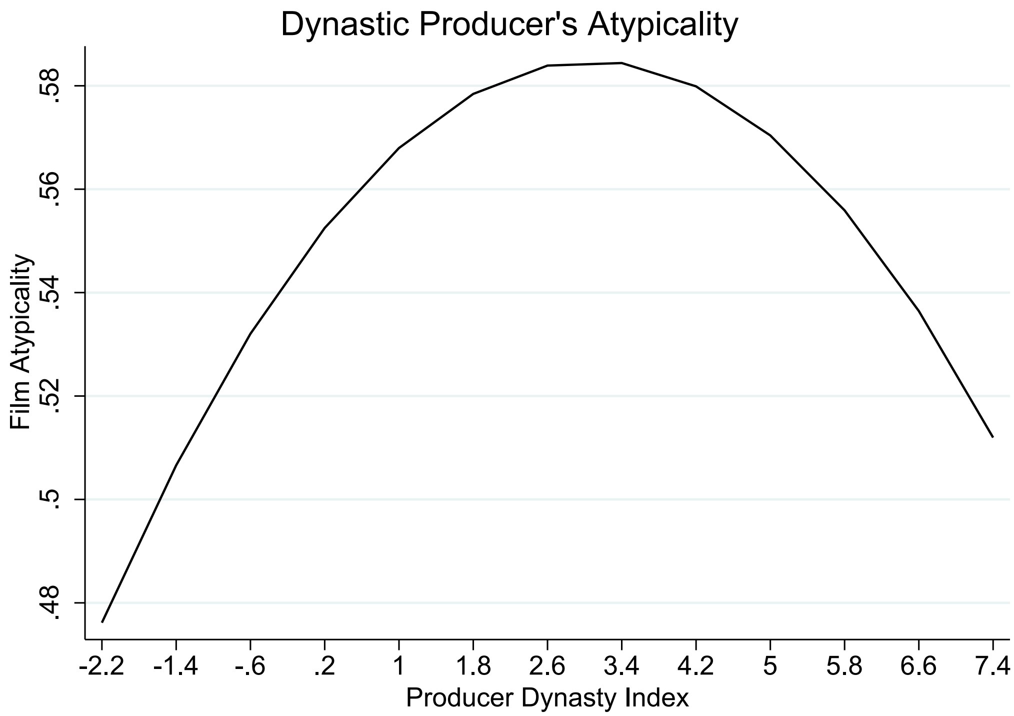
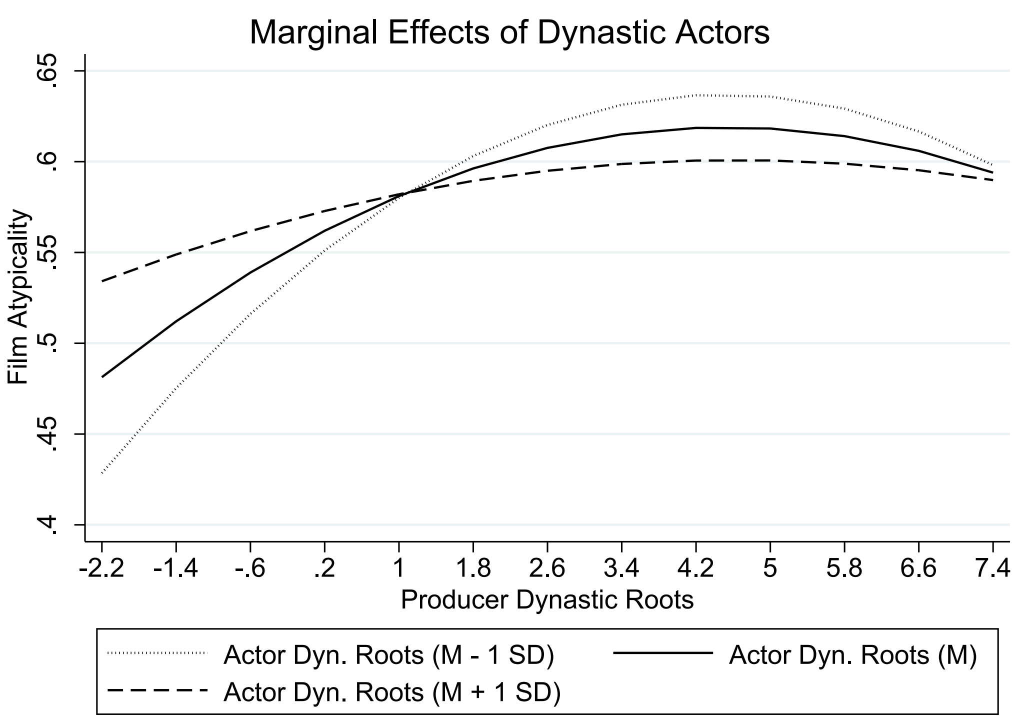


FIGURE 3



CONCLUSION

Innovation has long been recognized as a critical driver of growth, competitive advantage, and overall organizational success. It is a complex and multidimensional concept studied across various branches of science, engineering, humanities, arts, and social sciences. Organizational innovation, in particular, has been a significant focus of research, with scholars exploring internal and external factors that affect innovation production, how innovation is managed, and its effect on organizational behavior and outcomes. Recent advancements in computer science technologies have enabled new methods to understand and measure innovation, leading to substantial research exploring the meanings embedded in organizational innovation artifacts.

This dissertation presented three empirical studies that leverage contextual richness and uses novel text-based measures and methodological advancements to generate new insights to inform innovation theory. Chapter one focused on the division of labor between academic research and industry research to assess the value of patented technology. By analyzing the language used in patent documents and scientific publications, the study explored how to synthesize innovation knowledge generated from academic and industrial endeavors. Chapter two explored the impact of perceived competition on the dissemination of innovation or R&D findings on a regulatory platform. The study investigated how institutional interventions modify the competitive space of the firms and affect their disclosure behavior. Finally, chapter three developed theory and tested the relationship between the dynastic roots of producers and creative productions, examining how dynastic embeddedness enables and constrains innovation production. Together, these studies provide valuable insights into the production, sharing, and synthesis of organizational innovation. Each study identified a puzzle, delineated contributions, and suggested future research directions.

In conclusion, this dissertation asserts that innovation is a critical concept in organizational studies, with significant implications for a firm's competitive advantage and performance in an increasingly changing environment. Moreover, contextual elements influence innovation and are critical for understanding the complex and dynamic environment in which innovation occurs. This dissertation adds to the literature on organizational innovation by presenting three empirical studies that leverage contextual richness, novel text-based measures, and methodological advancements to generate new insights into the factors affecting innovation production, sharing, and synthesis. The studies provide valuable insights for scholars and practitioners and suggest ways for future research in the field of innovation management.

CONCLUSIÓN

Hace tiempo que se reconoce que la innovación es un motor esencial del crecimiento, la ventaja competitiva y el éxito general de las organizaciones. Se trata de un concepto complejo y multidimensional que se estudia en diversas ramas de la ciencia, la ingeniería, las humanidades, las artes y las ciencias sociales. La innovación organizativa, en particular, ha sido un importante foco de investigación, con estudiosos que exploran los factores internos y externos que afectan a la producción de innovación, cómo se gestiona la innovación y su efecto en el comportamiento y los resultados de la organización. Los recientes avances en las tecnologías informáticas han hecho posibles nuevos métodos para comprender y medir la innovación, lo que ha dado lugar a importantes investigaciones sobre los significados que encierran los artefactos de innovación organizativa.

En esta tesis se presentan tres estudios empíricos que aprovechan la riqueza contextual y utilizan nuevas medidas basadas en textos y avances metodológicos para generar nuevos conocimientos que sirvan de base a la teoría de la innovación. El capítulo uno se centra en la división del trabajo entre la investigación académica y la investigación industrial para evaluar el valor de la tecnología patentada. Mediante el análisis del lenguaje utilizado en los documentos de patentes y las publicaciones científicas, el estudio exploró cómo sintetizar los conocimientos de innovación generados a partir de los esfuerzos académicos e industriales. El capítulo dos exploró el impacto de la competencia percibida en la difusión de los hallazgos de innovación o I+D en una plataforma reguladora. Se investigó cómo las intervenciones institucionales modifican el espacio competitivo de las empresas y afectan a su comportamiento de divulgación. Por último, en el capítulo tres se desarrolló la teoría y se probó la relación entre el arraigo dinástico de los productores y las producciones creativas, examinando cómo el arraigo dinástico permite y limita la producción de innovación. En conjunto, estos estudios aportan valiosos conocimientos sobre la producción, el intercambio y la síntesis de la innovación organizativa. Cada estudio ha identificado un enigma, ha delineado las contribuciones y ha sugerido futuras direcciones de investigación.

En conclusión, esta tesis afirma que la innovación es un concepto crítico en los estudios organizativos, con implicaciones significativas para la ventaja competitiva y el rendimiento de una empresa en un entorno cada vez más cambiante. Además, los elementos contextuales influyen en la innovación y son fundamentales para comprender el entorno complejo y dinámico en el que se produce la innovación. Esta tesis se suma a la literatura sobre innovación organizativa presentando tres estudios empíricos que aprovechan la riqueza contextual, novedosas medidas basadas en textos y avances metodológicos para generar nuevos conocimientos sobre los factores que afectan a la producción, el intercambio y la síntesis de la innovación. Los estudios aportan información valiosa para estudiosos y profesionales y sugieren vías para futuras investigaciones en el campo de la gestión de la innovación.