Essays on Data-Driven Product Innovation in Organizations

HARVARD UNIVERSITY
Graduate School of Arts and Sciences

DISSERTATION ACCEPTANCE CERTIFICATE

The undersigned, appointed by the Committee for the PhD in Business Administration have examined a dissertation entitled

Essays on Data-Driven Product Innovation in Organizations

Presented by Ryan Thomas Allen

candidate for the degree of Doctor of Philosophy and hereby certify that it is worthy of acceptance.

Signature ______________________

Rory McDonald, Chair

Signature ______________________

Prithwiraj Choudhury

Signature ______________________

Gary Pisano

Date: March 30, 2023

Essays on Data-Driven Product Innovation in Organizations

Ryan T. Allen

Technology and Operations Management Department
Harvard Business School
Harvard University

Title Page

A dissertation submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Business Administration

May 2, 2023

Dissertation Committee:
Rory McDonald (Chair), Harvard Business School
Prithwiraj Choudhury, Harvard Business School
Gary Pisano, Harvard Business School

Required Disclaimers: *Researcher(s)' own analyses calculated (or derived) based in part on data from Nielsen Consumer LLC and marketing databases provided through the NielsenIQ Datasets at the Kilts Center for Marketing Data Center at The University of Chicago Booth School of Business. The conclusions drawn from the NielsenIQ data are those of the researcher(s) and do not reflect the views of NielsenIQ. NielsenIQ is not responsible for, had no role in, and was not involved in analyzing and preparing the results reported herein.*

I dedicate this dissertation to my wife Rose, who got me through it

Dissertation Advisor:
Rory McDonald

Author:
Ryan Allen

Essays on Data-Driven Product Innovation in Organizations

Abstract

How do organizations anticipate market demand to produce commercially successful product innovations? Much previous research emphasizes the advantages of being "data-driven"—utilizing data analytics or experimentation to validate market demand for new products. This dissertation investigates when and why such methods can be beneficial, as well as the circumstances in which they may limit innovation within organizations. In three empirical essays, I employ various theoretical lenses, such as behavioral strategy, decision-making, and organizational theory, to develop new theories on how and why data-driven methods systematically affect innovators' perceptions about the likelihood of success for their innovations.

The first essay employs a unique dataset comprising product-level sales data and employee résumé data from the consumer product sector, demonstrating that the efficacy of a data-driven approach in fostering innovation depends on the organization's methodological pluralism—specifically, the balance between quantitative and qualitative methods. The second essay elucidates one underlying reason why demand for new innovations can be challenging to predict using data-driven techniques. I create an agent-based model to demonstrate how diffusion dynamics obscure the observable demand for novel products, and I empirically validate the model using the same product-level sales data. The third essay presents an entirely different reason that data-driven decisions can mislead innovation—using data from developmental PC game development, I show that early enthusiastic users may provide demand signals unrepresentative of the broader market.

A recurring theme throughout the dissertation is a unique perspective on how organizations achieve competitive advantage in innovation: by shaping the methods their members use to perceive opportunities.

Acknowledgements

I tend to view the past through rose-colored glasses. So, partly for my own future benefit, I want to write it in stone that completing my PhD and finishing this dissertation was not all fun and games. In fact, it was incredibly challenging. The difficulty arose not only from the immense task of learning enough about a field to contribute to research or from the countless hours spent on what many might deem mundane data collection (which I enjoyed!). Rather, it was primarily challenging because, as is often the case among PhD students, I so often faced doubts about my choices and abilities, and uncertainty about the future. This was a process I simply could not have navigated without the support of my mentors, friends, and family.

Mentors

With this in mind, a few mentors deserve specific acknowledgements. First, I want to thank my main advisor, Rory McDonald. Rory has been irrationally dedicated to my success and well-being. For some advisors, I assume there is a certain *quid pro quo* transactional logic to taking a PhD student: "you help with my work, I help you finish your PhD." But Rory's commitment to me has been far less calculating, and far more generous. I know he has given much more than he has received. Throughout literally hundreds of meetings or calls, he has given me careful guidance on how to make a meaningful contribution to scholarly work, on the minutia of writing papers, on how to integrate into the scholarly community, and how to live a meaningful life. He did not shy away from critical feedback, but it was always given from a place of wanting me to succeed. He encouraged me to take unconventional leaps in my research, and then gave me enough support to actually make it happen. I consider Rory a lifelong mentor and friend. Thank you, Rory.

I am also grateful to the other members of my committee, who exhibited incredible generosity in their mentorship. I first started working with Raj Choudhury even before I officially

arrived as a PhD in Boston. From the outset, he trusted me with the co-author level responsibilities of analyzing data and writing papers. With Raj, there was no gradual transition—it was straight into the deep end. Over numerous meetings in his office, he offered practical advice on organizing my workflow and getting our papers published. Much of my ability to publish papers is directly attributable to Raj's mentorship. I am also grateful for his constant encouragement to "go out in the field" (including our trip to Shenzhen China) and to "swing for the fences." I am confident we will continue to collaborate on great work. Thank you, Raj.

I extend my gratitude to Gary Pisano as well. Although he formally joined my committee later than Rory and Raj, he was among the first professors I spoke with as a new PhD student. His work, reading recommendations, and thought processes significantly influenced the type of work I wanted to pursue. This impact is evident in this dissertation's topics and in my invitation to Gary to join the committee. He possessed an uncanny ability to quickly grasp the most critical issues and intriguing aspects of my work, providing insights that had eluded me for months. Much of this dissertation's most interesting content is a direct result of Gary's influence.

Other mentors at HBS faculty deserve my thanks. Thank you to Clayton Christensen for inspiring me and encouraging me early on in the program, and thank you to Tarun Khanna for early mentorship and collaboration.

I also want to thank my past academic mentors at BYU. Thank you to Professor Arden Pope, who treated me like a PhD student when I was a sophomore undergrad. The hours we spent coding SAS and writing drafts together in his office taught me to love research. Thank you to Professor Jeff Dyer, who sold me on the field of Strategic Management, showed me the way in, and has continued to support me. And thank you to Professors Shad Morris, Jim Oldroyd, and Ben Lewis, for showing me the ropes and sparking my interest in our field.

Friends

Many friends in the HBS program and the broader management field also deserve acknowledgement. While I had outstanding mentorship, I also relied on the day-to-day support of these friends. First, I want to thank my cohort-mates: Ashley Palmarozzo, Kala Viswanathan, MoonSoo Choi, and Ehsan Valavi. I appreciate your friendship and your support early on in the program, especially when Rose and I had our twins in our first year and I desperately needed help with classes. I also want to thank Aticus Peterson, and my other TOM friends, including Sourobh, Tommy, Daniel, Caleb, Paige, Natalie, and Maya.

I am grateful to the "Kool Krew" for keeping me laughing: Evan Defilippis, A Jay Holmgren, Emilie Aguirre, Jeff Steiner, and Ahmmad Brown. Let us always remember and "learn to see" that "the real treasure was the fond memories [we] were creating."

Thank you to the "Inc.lings" Jon Palmer and Kramer Quist, who were not only the best of friends but also pushed me to pursue interesting questions rather than just convenient ones. I will always cherish our weekly discussions over seltzer and carne asada burritos, and our Super PAC reunions with our families. Of course, I will always remember because we will continue these traditions for the rest of our lives.

I also want to extend my gratitude to an exceptionally friendly job market cohort for making the job market experience more, rather than less, pleasant. Special thanks go to Reuben Hurst, Anavir Shermon, Jacob Brown, Michael Impink, Amisha Miller, Nataliya Langburd-Wright, and Natalie Burford.

Family

My parents, Doug and Debbie Allen, deserve immense gratitude. They have been exceptionally supportive throughout my life. At home, they fostered an environment of learning, laughter, and love that allowed me to thrive. Now as a young parent myself, I am beginning to understand just how much they have done for me. They were also very attentive to my needs and the needs of my family during my PhD program. Thank you, mom and dad. And thanks to my brothers Jordan, Nate and Dylan for maintaining our strong friendships throughout the PhD despite the distance, and for always being interested in my work.

Last but most importantly, my wife and children deserve my deepest gratitude. When my wife Rose and I moved from Utah to Boston in August 2017, we had no children. Now we have four. Thank you to my children Charlie, Robyn, Elliot, and Gabe for keeping me grounded and reminding me what matters. PhD life, along with raising our family, was challenging at times. I am profoundly grateful to Rose for her love and patience. I dedicate this dissertation to her. We made it through the last six years together, and the experience has been a refiner's fire that brought us closer. Rose is my best friend and has kept me sane with her wit, sense of humor, no-nonsense attitude, and unwavering support. She is my most trusted advisor because she always says it like it is. Above anyone else I have acknowledged, Rose deserves significant credit for the completion of this dissertation. Thank you.

Ryan Allen
Cambridge, MA
May 2, 2023

Table of Contents

Introduction

How do organizations anticipate market demand to produce commercially successful product innovations? Commercializing innovations is fraught with uncertainty, and most new products fail (Schneider & Hall, 2011). An influential perspective in the strategy literature is that organizations may achieve a higher rate of success by being "data-driven"—relying on data analytics or experimentation to quantitatively validate demand for new products. The argument is that data-driven organizations can process more information about customer preferences (Brynjolfsson, Jin, & McElheran, 2021; Wu, Lou, & Hitt, 2019), make more objective decisions (Brynjolfsson & Mcelheran, 2019; Glaeser, Hillis, Kim, Kominers, & Luca, 2021), and validate innovation opportunities with scientific rigor (Camuffo, Cordova, Gambardella, & Spina, 2019; Koning, Hasan, & Chatterji, 2022; Levinthal, 2017)—all of which would, theoretically, augment organizations' ability to select the most promising innovations.

But organizational scholars have called this view into question, highlighting that a data-driven approach can negatively impact innovation efforts. According to this view, the most commercially successful innovations are relatively difficult to quantitatively validate, so data-driven selection processes may divert resources from high potential innovation opportunities (Baldwin & Clark, 1994; Benner & Tushman, 2002; Can Deniz, 2020; Christensen & Bower, 1996; Felin, Gambardella, Stern, & Zenger, 2019; Ghosh, 2021).

Recognizing the tensions inherent in these perspectives, my dissertation aims to develop deeper theoretical understanding of when and why data-driven organizations attain innovative success (or failure). As my primary theoretical lens, I engage with the behavioral strategy literature (Augier, Fang, & Rindova, 2018; Gavetti, 2012). This emerging school of thought views behavioral processes (e.g., managerial cognition or organizational culture) as a key and underappreciated source of sustained superior performance in organizations. In this spirit, I explore how managerial cognition around data-driven methods, both on an individual and on an organizational level, shape innovation outcomes in various circumstances. I explore these themes in three chapters.

In chapter 1, I ask whether organizations with data-driven cultures are more innovative. I show that this depends on what I call the methodological pluralism of organizational culture—the extent to which a range of different methods are valued in the organization. Using data on 3,500 product launches and text from employee résumés at 61 large consumer product firms, I show that increasing quantitative analysis decreases innovation performance when qualitative analysis is low, and, conversely, increases when qualitative analysis is high. A culture dominated by quantitative analysis is especially disenabling when launching products that are novel, or when targeting markets characterized by higher uncertainty. The paper highlights that reliance on quantitative data does not stifle innovation, but exclusive reliance does.

In chapter 2, I use an agent-based model to explain an underlying mechanism for why demand for new innovations can be difficult to anticipate using data-driven methods. Specifically, the paper seeks to explain what I call the "market size inversion" puzzle: the observation that many breakthrough products achieve success despite low initial expectations, while other products commonly fail despite high projected demand. I develop and elaborate in explanation of this puzzle as a function of social diffusion processes. Relatively novel products are more ambiguous to evaluate, so potential adopters rely more on neighbors' endorsements in their adoption decision. Therefore, a larger portion of demand does not exist until after the product diffuses, and initial market size estimates for relatively novel products will tend to be downward biased. The model also provides an alternative (or complementary?) driver for what I observed in chapter one:

because novel products appear to have lower market demand than non-novel products with larger anticipated market sizes, organizations that rely exclusively on quantitative estimates of market size will launch fewer highly novel or breakthrough products. I empirically validate the model using sales and attribute data from 1,600 consumer product launches, combined with measures of firms' use of quantitative market sizing analysis. The paper provides a demand-side framework to complement the existing competition-based strategy explanation for the same puzzle—that breakthroughs are often surprising because if they were not, the opportunity would have been competed away.

Finally, in chapter 3, I shift my focus to an entirely different reason that data-driven decisions can mislead innovation: early users that are unrepresentative of the broader market. Whereas prior research has largely focused on the benefits of leveraging user communities as a resource for experimentation and feedback in product development, this paper posits a boundary condition: reaping these benefits is contingent on the degree to which the user community accurately represents demand in the broader market. Using longitudinal and interview data on experimental PC game development, I find that adapting games in response to early user-community feedback increases games' growth in the market, on average. But this effect is reversed when the feedback comes from a homogenous user community that is concentrated in a narrow market segment. Such homogenous communities can produce signals of market demand that, when incorporated into the game, diminish its appeal to potential customers outside the community.

A theme throughout the dissertation is a novel perspective on how organizations attain competitive advantage in innovation: through shaping the epistemological beliefs of organizational members. What people believe about the methods they use profoundly shapes the opportunities they see, the decisions they make, and therefore the outcomes of the innovations they pursue.

Chapter 1. Methodological Pluralism and Innovation in Data-Driven Organizational Cultures

1. Introduction

In 2012 Procter & Gamble (P&G) launched Tide Pods, a product that reshaped the sleepy laundry category and generated billions in revenue. Surprisingly, P&G's disciplined, "data-driven" culture had nearly rejected the project, whose "unimpressive" projected returns were based on a mix of quantitative consumer tests and past product performance data. But a handful of senior managers persisted in pushing for the launch, bolstered by a contradictory qualitative insight: "[consumers] would hold it in their hands *like it was a jewel*".[1] Such qualitative observations prompted them to question their standard quantitative evaluations, to conduct additional analyses, and ultimately to prepare for a large-scale launch.

The Tide Pods vignette tidily frames the existing literature's mixed perspectives on how data-driven organizational cultures—in which new-product investment decisions are based on empirical evidence of market demand—influence new-product innovation. One widespread belief is that organizations that use more data analytics and experimentation excel at commercializing innovations because they are better equipped to test their assumptions and validate consumer preferences (Camuffo et al., 2019; Koning et al., 2022), and may be less subject to confirmation bias and internal politics (Brynjolfsson & Mcelheran, 2019; Brynjolfsson & McElheran, 2016). Anecdotes like the Tide Pods case—successful innovations unsupported by the data—are explained as outliers, or as a result of unsophisticated data analysis.

[1] Quotations are from the author's interviews with company executives.

In contrast, a skeptical stream of innovation research by organizational scholars suggests that more data-driven decision-making can discourage innovation. Proponents of this view assert that, due to the inherent uncertainty of innovation, the most commercially successful innovations are relatively difficult to validate *ex ante* using quantitative measurements. Thus data-driven organizations, which favor measurable projects, tend to divert resources away from the highest potential innovation opportunities (Baldwin & Clark, 1994; Benner & Tushman, 2002; Can Deniz, 2020; Christensen & Bower, 1996; Felin et al., 2019; Ghosh, 2021). This research stream views cases like Tide Pods as the norm rather than the exception, and we might infer that organizations seeking to innovate would benefit from reducing reliance on quantitative analysis. In recognition of the tensions between these perspectives, I ask: under what circumstances are data-driven organizations more innovative?

Note that embedded in the Tide Pods vignette is an underappreciated consideration common to both viewpoints. P&G did not select the successful product by increasing, nor by eliminating, quantitative analysis of consumer data. Rather, managers recognized the product's potential because the qualitative evidence of consumer delight was viewed as legitimate enough to instigate a challenge to the seemingly contradictory quantitative evidence. The organizational culture at P&G thus exhibited a high level of *methodological pluralism*—valuing a range of different methodologies—for evaluating new product potential.

This paper argues that the methodological pluralism of an organizational culture is a crucial contingency that accounts for whether a data-driven organization is more, rather than less, innovative. This claim is built on a more multi-dimensional view of what it means to have a "data-driven organizational culture". Drawing on recent advances in the organizational-culture literature (Chatman & O'Reilly, 2016), I distinguish the *magnitude* of an organization's use of quantitative analysis from its methodological *pluralism*. The distinction acknowledges that an organization can make heavy use of quantitative analysis (e.g., data analytics) without dogmatically excluding other types of analysis (e.g., customer interviews) in evaluating potential innovations. Some research on innovation implicitly conflates these two dimensions (by assuming that high magnitude implies low pluralism); meanwhile proponents of data-driven decision-making have overlooked the potential detrimental effects of low methodological pluralism.

Drawing from previous research on innovation strategy and organization theory, I theorize that organizations with high methodological pluralism are more likely to produce products that have the potential to become breakthrough commercial successes. In this paper, I specifically conceptualize methodological pluralism as valuing both quantitative and qualitative analysis. Such pluralism allows a more multi-dimensional evaluation than quantitative analysis alone, thereby nullifying several cognitive and organizational mechanisms that can downward bias purely quantitative estimates of demand for relatively novel products. But quantitative analysis also provides useful additional information that balances out the blind spots of purely qualitative analysis. Thus, contrary to much prior innovation research, I argue that organizations that use more quantitative analysis in new-product innovation will actually produce *more* commercially successful innovations—but only if their cultures are pluralistic enough to rely heavily on qualitative analysis as well.

To test this theory, I constructed a unique dataset that tracks the commercial performance of 3,502 new products launched by 61 large consumer-packaged-goods (CPG) firms between 2010 and 2016. Using a rich data source—excerpts from employees' résumés describing their innovation-related work—I use natural language processing to determine how much each organizational culture uses qualitative and quantitative analysis in the innovation process.

Combining these measures with data on new-product sales, I confirm that an organization's use of qualitative analysis positively moderates the relationship between its use of quantitative analysis and the commercial performance of new-product innovations.

I also validate the theoretical underpinnings of my findings with additional mixed-methods analyses, leveraging archival analysis and interviews of CPG industry professionals. First, I show that quantitative analysis is associated with producing fewer highly novel products, and that organizations with low methodological pluralism produce fewer breakthrough successes when launching novel products. Second, I explore the effect of methodological pluralism within, rather than between, organizational members. Third, I confirm market uncertainty as an important boundary condition: the theory has more explanatory power in conditions of high market uncertainty. And finally, I investigate the antecedents of data-driven culture: management fads, not organizational learning, appear to account for excessively quantitative cultures.

This study contributes to research on innovation strategy in organizations in several ways. First, prior innovation research suggests that data-driven organizations produce incremental rather than breakthrough innovations (Benner & Tushman, 2002; Can Deniz, 2020; Christensen & Bower, 1996). I clarify that this is only likely to be true when quantitative analysis is used to the exclusion of other methods. In fact, in my sample, the organizations that produce the most breakthroughs use the *most* quantitative analysis—provided that they used qualitative analysis liberally as well. Thus, an underappreciated and underlying driver of a data-driven organization's failure to produce breakthroughs is not its use of quantitative analysis *per se*, but a lack of methodological pluralism. Second, research on strategic decision-making has consistently demonstrated that data-driven models outperform humans in repeated-judgment contexts (Allen & Choudhury, 2022; Glaeser et al., 2021; Hoffman et al., 2018; Kleinberg et al., 2018). Against this backdrop, I identify product innovation as a context that produces a theory-refining anomaly: a context in which success is characterized by questioning and interrogating the data with other sources of information—perhaps indicative of a more theory-driven entrepreneurial approach (Camuffo et al., 2019; Felin et al., 2019). Finally, this study contributes to an emerging body of literature on the link between organizational culture and strategy, and highlights a new way that culture impacts strategy: by shaping how organizational members perceive potential opportunities. Applying this cultural lens to studies of strategic innovation is a perspective that has repeatedly been called for (Felin & Zenger, 2017; Gavetti, 2012; H. Li & van den Steen, 2019; Puranam, 2018) but seldom applied in empirical studies.

2. Theoretical Background

Two research streams inform the link between data-driven culture and innovation in organizations. One stream, at the intersection of the economics-of-IT and strategy literature, advances the view that data-driven decision-making promotes innovation. The argument is that organizations that use more data analytics are better equipped to test their assumptions about the market (Camuffo et al., 2019; Koning et al., 2022), process larger volumes of external information, such as customer preferences (Brynjolfsson et al., 2021; Hitt et al., 2015; Müller et al., 2018; Tambe, 2014; Wu, Hitt, et al., 2019), and may be less prone to internal organizational politics or confirmation bias that support management's pet projects (Brynjolfsson et al., 2011; Brynjolfsson & Mcelheran, 2019; Glaeser et al., 2021)—all of which would, theoretically, augment organizations' ability to produce the most promising innovations.

This viewpoint is consistent with empirical evidence that, on average, more data-driven organizations enjoy higher sales, higher productivity, and higher market value (Brynjolfsson &

Mcelheran, 2019; Brynjolfsson & McElheran, 2016; Conti et al., 2020; Koning et al., 2022; McAfee & Brynjolfsson, 2012; Müller et al., 2018; Tambe, 2014). These benefits do not accrue equally, however, and depend on the presence of certain organizational features (Brynjolfsson et al., 2021; Koning et al., 2022; Müller et al., 2018; Tambe, 2014; Wu, Lou, et al., 2019) and on the type of innovation under consideration. With regard to the latter, a recent study that examines the link between data analytics adoption, patents and sales suggests that data analytics contribute most to activities that are more amenable to quantitative analysis, such as process-oriented or incremental innovation (Wu, Hitt, et al., 2019). Innovations based on novel technologies, the authors argue, benefit less from data analytics because existing data is most useful for improving existing products and processes.

Overall, empirical evidence from this research stream appears consistent with the idea that data-driven organizations are more innovative, with the caveat that data is less useful in certain contexts. Mcafee and Brynjolfsson (2012) neatly summarize the view: “Data-driven decisions are better decisions—it’s as simple as that.”

Another research stream, built on organizational theories of innovation, promotes a different view: increasing data-driven decision-making in organizations may actually suppress innovation. This perspective has deep roots in classic work on the “productivity dilemma” originally observed at automobile-manufacturing firms, whose relentless focus on productivity gains inhibited the ability to innovate and adapt (Abernathy, 1976). Picking up this theme, subsequent work warned against allocating innovation resources based on what was measurable in the short term, and thus unintentionally tipping the balance too far toward incremental exploitation rather than innovative exploration (Ahuja & Lampert, 2001; Benner & Tushman, 2003; Christensen & Bower, 1996; Leonard-Barton, 1992; March, 1991; O’Reilly & Tushman, 2008).

For example, Baldwin and Clark (1994) argue that, when U.S. firms adopted capital budgeting systems, “managers at all levels lacked objective data and analytic tools” to invest in less-measurable investments (Baldwin & Clark, 1994); certain organizational capabilities were ignored, to the long-term detriment of these firms’ new-product innovation and growth. Similar mechanisms were observed at firms that adopted ISO 9000 process management, where researchers observed that “exploratory activities [were] increasingly unattractive compared with the short-term measurable improvements,” and that exploratory patents were thus “crowded out” by exploitative patents (Benner & Tushman, 2002, 2003). In sum, these studies established that focusing on quantitatively measurable outcomes unintentionally inhibits organizations’ investment in less-measurable exploratory innovations.

In a similar vein, recent studies have theorized that innovating based on quantitative evidence of consumer preferences may lead to incremental, rather than radical, innovation (Felin et al., 2019). These studies have focused on organizations’ use of A/B testing, arguably the gold standard among quantitative tests of consumer preferences. For instance, Ghosh (2021) found that inexpensive access to A/B testing tools in organizations had the unintended consequence of shifting engineering and cognitive resources away from intentional planning, resulting in less strategic and impactful experiments. Another study found that U.S. newspapers that adopted A/B testing tools were likely to make incremental, rather than radical, changes to their websites (Can Deniz, 2020).

Thus, despite shifts over time in the specific tools (from ISO 9000 to A/B testing), sources of uncertainty (from technical uncertainty to market uncertainty), and types of innovation (from process to product) being studied, this research stream has held to the view that data-driven decision-making has the potential to discourage radical or breakthrough innovations in organizations. The underlying argument is that, in intra-organizational competitions with other projects for scarce resources (Bower, 1970; Burgelman, 1991; Choudhury, 2017; Christensen & Bower, 1996; Klingebiel & Rammer, 2014), data-driven organizations tend to select relatively measurable—and thus relatively incremental—innovations.

Taken together, these two research streams suggest a tradeoff: data-driven organizational cultures excel at commercializing incremental innovations, but at the cost of unintentionally deprioritizing less measurable radical or breakthrough innovations. Although insightful, this view is incomplete. Prior lines of inquiry examine quantitative analysis in a vacuum, independent of other types of analysis. These examinations rely on a single conceptual lever: the magnitude of quantitative analysis. This one-dimensional view implicitly equates a higher magnitude of quantitative analysis with more strict compliance with quantitative output—a foundational assumption of the proposed data-driven incremental/breakthrough tradeoff. But would the tradeoff hold if quantitative analysis represented just one tool in an organization's toolbox? Drawing on recent conceptual advances in the organizational-culture literature (Chatman & O'Reilly, 2016), I question the proposed incremental/breakthrough tradeoff by taking a more multi-dimensional view of "data-driven organizational culture". I propose that a more complete understanding of the relationship between a data-driven organization and innovation requires a broader view of the culture around using a variety of tools and methods.

Organizational culture is often understood as a system of collectively accepted assumptions, beliefs, and norms (Schein, 2010), to which members internalize and conform through socialization (Goldberg et al., 2016; Schein, 2010; Srivastava et al., 2018; Van Maanen & Schein, 1977). Recent developments in the literature have emphasized that these collective norms can be parsed into three distinct dimensions: (1) the content of the norms that are viewed as important;

(2) the cultural consensus, or whether everyone shares the same norms; and (3) the intensity of beliefs about the norm (Chatman & O'Reilly, 2016). Separating these dimensions has been useful for establishing, for example, that cultural consensus can affect performance regardless of the content of the norms (Corritore et al., 2020; Sørensen, 2002), or that specific norm content (e.g., a norm for adaptability) can make an otherwise rigid high-consensus culture more adaptable (Chatman et al., 2014). Follow-on work has further refined these dimensions, demonstrating different outcomes for organizations with high intra-personal rather than inter-personal cultural consensus (Corritore et al., 2020; Marchetti, 2019). In this paper, I primarily focus on norm content—specifically, which methods are viewed as legitimate, and the degree to which different types of methods are tolerated. (In the Discussion section, I consider how my work represents a further refinement to the "content" dimension of culture by focusing on the multi-dimensionality of the norms themselves.)

An organizational culture in which different types of evidence are accepted as legitimate methods of verifying assertions—in this case, assertions about potential market demand for new products—has what I call high methodological pluralism. Although many organizations favor the rational calculations of quantitative analysis (Feldman & March, 1981; Kellogg et al., 2020), this is far from a universal constant. Comparisons of scientific cultures indicate that different communities vary in the degree to which they embrace methodological pluralism (Knorr-Cetina, 1999; Lamont, 2009). For instance, Akerlof (2020) observed that economists favor "hard" methods so wholeheartedly that researchers ignore topics that are difficult to represent in a "hard" way; meanwhile, other disciplines—such as political science and sociology—have traditionally accepted a wider range of tools and methodologies (Lamont, 2009). Organizations, like scientific communities and academic disciplines, may also vary in their degree of methodological pluralism

(Anthony, 2018; Kaplan, 2011; Rindova & Courtney, 2020), but its impact on innovation outcomes remains unexamined.

This paper specifically conceptualizes methodological pluralism as the degree to which organizational members use both quantitative and qualitative analysis to evaluate the market potential of new products. I define as *quantitative* those methods or tools that analyze precise and aggregable numerical measurements (e.g., data analytics, fixed-response surveys, or market testing); *qualitative* methods or tools analyze non-numerical observations (e.g., focus groups, ethnography, or interviews). There are other possible ways to capture methodological pluralism; I focus on these two broad methodological categories for two reasons. First, qualitative and quantitative analysis are arguably the two most relevant categories of empirical methods available to innovators exploring and validating market demand for new products (Martin & Golsby-Smith, 2017; Seemann, 2012; Shapiro, 2018). Second, because the fundamental assumptions of quantitative and qualitative analysis are rooted in different ontological and epistemic paradigms,[2] their coexistence in the same culture is a potential source of tension (Lamont, 2009). For instance, many of the most useful qualitative insights could easily be characterized as cherry-picking—precisely what the quantitative paradigm aims to guard against (Siggelkow, 2007). As a natural foil for quantitative analysis, the coexistence with qualitative methods offers a compelling indicator of methodological pluralism.

In sum, a more comprehensive understanding of the link between data-driven culture and innovation requires a deeper understanding of what is meant by the term "data-driven culture." I distinguish the magnitude of an organization's use of quantitative analysis from its methodological

[2]Quantitative analyses tend to operate according to positivist assumptions (i.e., reality is best understood by means of empirical measurement of a single factual reality); qualitative analyses tend to rely on a constructivist or interpretivist paradigm (i.e., reality is best understood by inquiring into people's experiences and interpretations of reality).

pluralism—conceptualized as its degree of openness to both quantitative and qualitative analysis. The next section posits a formal hypothesis.

3. Hypothesis

How Quantitative Analysis Can Suppress Innovation: Cognitive and Organizational Mechanisms

Two interrelated mechanisms explain why the use of quantitatively data-driven analysis alone can impede an organization's ability to produce high-potential new-product innovations. The first is a cognitive mechanism: organizational members in data-driven cultures may unquestioningly accept assumptions embedded in quantitative estimates that systematically undervalue innovations with high potential for commercial success. Assumptions embedded in tools and methods shape users' perceptions, highlighting some dimensions of reality while obscuring others (Beunza & Stark, 2004; Kaplan, 2011; Orlikowski, 1992, 2007). Though not always salient to users, these assumptions may be accepted without question in some organizational cultures (Anthony, 2018). For example, Anthony (2021) has documented that many junior investment bankers unquestioningly accept the quantitative metrics embedded in CapIQ and FactSet without realizing how those assumptions have shaped their industry valuations. Assumptions embedded in methods and tools can readily shape the processes that determine how time, attention, and financial resources are allocated. For instance, Can Deniz (2020) has shown that developers who based their web-design decisions on the results of A/B tests assumed that what the test measured (an immediate bump in web traffic) aligned with the desired outcome (long-term value). Whether the developers' assumption was conscious (e.g., they optimized for measurable results to boost their careers) or unconscious (e.g., they didn't realize how A/B testing shaped their decisions), it promoted merely incremental changes in web design.

Unquestioning acceptance of assumptions embedded in quantitative analysis may be particularly problematic because novel products with high commercial potential may tend to be systematically undervalued in quantitative market sizing (Allen, 2022). The strategy literature asserts that the commercial opportunities with the greatest potential reward are novel and unfamiliar, and thus less likely to be perceived or contested (Denrell et al., 2003; Gavetti, 2012; Gavetti & Levinthal, 2000; Levinthal, 1997; Levinthal & March, 1993; McDonald & Allen, 2021). In fact, many valuable innovations become successful precisely because they defy current consumer preferences, trends, and patterns (Allen, 2022; Felin & Zenger, 2017). Thus, accepting the assumptions embedded in quantitative estimates of market size can systematically underestimate potential commercial breakthroughs, making them appear less attractive than other innovations.

A second mechanism operates at the organizational level. The highest-potential product innovations may be the most difficult to measure quantitatively because they are likely to be characterized by high levels of novelty and of market uncertainty—that is, unmeasurable and unquantifiable risk about demand for a product (Knight, 1921) that cannot be entirely resolved by testing or market research prior to launching the innovation (Bhide, 2003; Gao & McDonald, 2022; Levinthal, 2007; McDonald & Eisenhardt, 2019; McDonald & Gao, 2019). If leaders view quantitative analysis as the sole legitimate evidence, organizational members will be unlikely to marshal the political will or take the career risk to champion quantitatively unmeasurable investments during the resource-allocation process (Bower, 1970; Burgelman, 1991; Gilbert & Bower, 2005), particularly if such metrics are reified by rituals of quantification (Mazmanian & Beckman, 2018). Especially in organizations that require more handoffs to approve new ideas (Berg & Yu, 2021), such organizational pressures may lead organizational members to cull, or even to refuse to consider, less measurable high-potential products in the innovation pipeline (Benner & Tushman, 2002; Can Deniz, 2020; Christensen & Bower, 1996; Vinokurova & Kapoor, 2020).

How Methodological Pluralism Helps Innovation: Cognitive and Organizational Mechanisms

In regards to both the cognitive and the organizational mechanism, it is less likely that increased use of quantitative analysis will undermine innovation performance in organizational cultures characterized by high methodological pluralism—those that also liberally use qualitative analysis. First, in such cultures, assumptions embedded in quantitative analysis are called into question by potentially conflicting perspectives from qualitative analysis. Qualitative and quantitative evidence often conflict: qualitative and quantitative methods can be seen as opposite ends of an epistemic spectrum (Jick, 1979). Quantitative methods are better suited to measuring reliable patterns in large populations, but they are inherently reductive. Because humans have limited cognitive information-processing abilities, this reductionism makes it possible to digest complexity into statistically representative patterns. But subtracting the complexity of context can obscure dimensions of value that are not already captured in the measurements, and can make it difficult to explain *why* certain patterns arise (Edmondson & Mcmanus, 2007). Qualitative tools, on the other hand, are richly descriptive and retain context, but may fail to capture representative patterns (Kaplan, 2016). These relative strengths favor better suited to exploring unconsidered dimensions of value, to understanding why patterns arise, and to observing edge cases (Siggelkow, 2007) that may be indicative of future untapped value (Gavetti et al., 2017; Rindova & Courtney, 2020). The strengths of the two types of evidence reinforce each other, much as people with diverse backgrounds can use their dissimilar perspectives to avoid groupthink (Choudhury & Haas, 2018; Lix et al., 2019; Singh & Fleming, 2010). Thus, organizations whose cultures are characterized by methodological pluralism will benefit from a process of validation that compensates for the blind spots of any single method (Kaplan, 2016; Page, 2018).

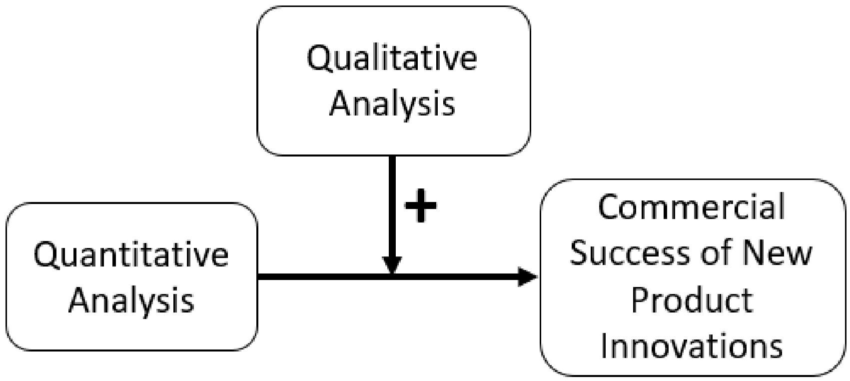
Second, at an organizational level, in methodologically pluralistic cultures, failure to procure quantitative evidence of market demand will be less likely to kill a new-product initiative. Because such quantitative evidence may be systematically harder to procure for breakthrough innovations (Allen, 2022; Felin & Zenger, 2017, 2018; Gavetti, 2012; Zellweger & Zenger, 2021), viewing qualitative evidence as a legitimate rationale for new-product decisions gives breakthroughs a better chance of surviving a competitive resource-allocation process. When qualitative and quantitative evidence do not align, organizational members may question the certainty of quantitative forecasts, and re-interpret the existing evaluation criteria in a way that is favorable to novel innovations (Vinokurova & Kapoor, 2020). For example, quantitative market forecasts for Tide Pods appeared to indicate tepid demand—but those low forecasts were called into question by qualitative observations of "consumer delight" in focus groups. This qualitative observation caused leaders to question the assumptions of the market tests, to conduct additional analyses, and ultimately to grant the project enough resources and enough political capital for a large-scale launch.

Finally, when complemented by qualitative analysis, pursuing more quantitative analysis may *improve* innovations' commercial performance. There are two reasons for this phenomenon. First, though quantitative market-size estimates for potential breakthroughs may tend to be downward-biased, on average, this is not a universal constant. Some highly successful products are readily discovered or validated by quantitative analysis of consumer preferences (McAfee & Brynjolfsson, 2012). Consumer tests and experimentation can reveal surprising opportunities that defy executive experience and conventional wisdom (McDonald & Eisenhardt, 2019; Thomke, 2020). This can in turn mobilize political support for a radical idea that would otherwise be difficult to sell internally. For example, Netflix moved into original content with its hit series "House of Cards" in part because data indicating strong viewer demand for the show (Patel, 2013) had enhanced its confidence. Second, relying solely on qualitative analysis would generate its own set of weaknesses and barriers to innovation. Quantitative analysis gives an organization more information about consumer preferences and about the market, enabling organizational members to dig deeper and to triangulate on a better estimate of the potential value of an innovation than qualitative analysis alone can provide. In short, organizations with pluralistic cultures may be well positioned to reap the benefits of quantitative analysis without incurring the accompanying costs.

To summarize, in organizations that rely solely on quantitative analysis (i.e., those with low methodological pluralism), commitment to selecting products in a data-driven way may become an obstacle to innovation. But in organizations that routinely use both quantitative and qualitative analysis (i.e., those with high methodological pluralism), quantitative analysis provides useful information and an additional lens for evaluating a potential innovation—another tool in the toolbox—leading to more accurate valuations of potential breakthroughs. Formally, I hypothesize:

> **Hypothesis:** *Organizations' use of qualitative analysis positively moderates the relationship between quantitative analysis and the commercial performance of new product innovations. Organizations that use high levels of both qualitative and quantitative analysis will launch the most successful new products.*

Figure 1. Hypothesis



4. Methods

Research Context

I test my hypothesis in the context of the U.S. consumer-packaged-goods (CPG) sector. The CPG sector consists of all firms that manufacture products sold to consumers via retail channels—a market with estimated U.S. sales of $815 billion in 2019. The sector spans a wide range of product categories, including food and beverage, household cleaning, beauty, over-the-counter drugs, and electronics.

Two characteristics of the CPG sector make it a fertile setting in which to test my hypothesis. First, the sector is characterized by high market uncertainty, and successful new-product innovations (especially breakthrough hits) are imperative for organizations' growth and survival. According to the Nielsen Retail Scanner dataset (introduced in the next section), the 5% of products with highest first-year sales accounted for 67% of all new-product first-year sales in 2016.

Second, in the CPG context, the culture around quantitative and qualitative methods is highly influential within the product-innovation process. As in many industries, the cost of developing and launching products is sufficiently high that they must be validated before being fully developed. To control costs, most CPG organizations require product concepts to pass through formal evaluation checkpoints; at these junctures, empirical evidence of consumer demand is central to determining which products will receive further investment. New-product initiatives are typically developed and validated throughout a 1–2-year innovation process. Most CPG organizations use a funnel or waterfall innovation process. The funnel is initially wide; it encompasses broad areas for potential innovations, pinpointed in macro trends (e.g., rising interest in DIY beauty solutions, or problems with sleep). In these general areas, organizational members use a variety of qualitative and quantitative methods to generate early-stage ideas, and to select among them. For example, early-stage concept tests compare consumers' reactions to product mock-ups and descriptions. A product progresses through a stage-gate process in which it must reach specific milestones before qualifying for more investment. For example, in the later stages of the process, in order to qualify for final market launch, a product may have to perform well in a market test. Throughout this process, an array of qualitative and quantitative analyses are used to validate demand for products and features: analysis of purchasing patterns in consumer-panel data, ethnography, consumer interviews, focus groups, concept tests, and market-size forecasting based on the success of past similar products. Thus, how the organizational culture shapes members' views of the legitimacy of qualitative and quantitative methods plays a key role in the screening of and investment in potential new products.

To augment my archival analysis, I conducted semi-structured interviews with 36 CPG industry informants. Consistent with similar multi-method investigations (Bermiss & McDonald, 2018; Pahnke et al., 2015), I used the snowball technique to identify suitable interviewees. I interviewed executives, R&D personnel, data scientists, and product managers at eight leading companies representing a broad range of product categories. My goal was to understand how each organizational culture evaluated new-product innovations, and to explore how these cultures develop. These field-based interviews provided useful contextual information about innovators' perceptions of how their organizations' epistemic cultures[3]—their norms and values around "how they know what they know"—shaped innovation.

One of the most salient observations from the interviews was that CPG professionals are readily able to characterize how different data-driven methods were viewed in their own organizational culture. For instance, one senior product manager described inter-organizational variation by recounting that a former colleague had had difficulty rallying support for his projects when he moved to a relatively data-driven organization: he "had a really hard time selling the organization on supporting him because he's not coming back with that rigor. . . . He came from . . . a place that didn't operate the same way." Another manager noted that being "data-driven" was part of his company's identity: "we consider ourselves to be a very data-driven company. I know that moniker has been taken up by many, many companies of late, but we feel like we have always been a data-driven company—and that is evident in all of our systems and processes."

A second salient observation is that interview participants believed an organization's epistemic culture to have significant consequences for innovation. One product manager recalled: "The ubiquity of data genuinely changed the way in which decisions were made—if there was data to back a decision, it was often approved. . . . Otherwise, it was often dismissed." A data scientist noted that less-measurable projects were an "uphill climb," and that product managers wouldn't champion them unless they were thoroughly convinced and had sufficient "political capital"; he added that he would even "actively discourage a junior PM doing a [less-measurable project]" due

[3] I borrow the term *epistemic culture* from Knorr-Cetina (1999). See also Kaplan (2011).

to career risks. Some interviewees recognized that this phenomenon was particularly problematic for potential breakthrough innovations, and chafed at excessively data-driven cultures. In one R&D manager's words:

> That was the insanity of data-driven decision making at [our company]. Because decision makers need data, data that have been validated, because we have done the same test for years, . . . [but] if I do the things that [are] already in the data, someone else will be able to catch up. That's my dilemma.

Others recognized the value of data in new-product innovation, while acknowledging that "getting that balance [in data-driven decision-making] right is a tension for an organization," in one manager's words.

Taken together, the field interviews reveal that CPG firms' epistemic cultures are salient to employees, that culture has tangible consequences for new-product innovation, and that achieving the right balance is a difficult tension. The Results and Discussion sections will return to these interviews to highlight additional themes and to complement my primary quantitative archival dataset.

Sample and Data

To formally test my hypothesis, I constructed a unique dataset that provides both a granular view of product-level attributes and sales, as well as an in-depth view of the methods most valued in the innovation process within each organization. The final dataset is the result of an extensive data-collection process involving merges of several product-level and organization-level datasets, matched to aggregated employee-level information to track 3,502 product launches at 61 large CPG organizations during the period 2010–2016.[4]

[4] My data extend through 2018. Because my measure of commercial performance uses sales data for 2 years after the initial launch, 2016 is the final year studied.

I constructed measures of product features and commercial performance using the Nielsen Retail Measurement Services (RMS) scanner dataset.[5,6] To avoid counting minor changes to existing products as new products, I aggregated UPC barcode-level data up to the brand/product-module level.[7] I matched products to firms using the GS1 UPC-company-matching database.

I constructed measures of each organization's use of quantitative and qualitative analyses using résumé job description text from a popular online career networking website (DeSantola et al., 2020).[8] Because my aim was to measure organizational culture as it pertained to product innovation, I extracted résumé text only for employees whose reported job functions included "product management" or "research," using the networking website's internal job-function classification system.[9] I obtained 101,919 unique employee résumés describing 182,403 positions[10] from 2010 to 2016 for the CPG firms in my sample. Although résumés are a novel data

[5] These data were accessed through the Kilts-Nielsen Data Center at the University of Chicago Booth School of Business.

[6] The Nielsen RMS data set is one of the most comprehensive and representative point-of-sale retail datasets available. It covers most sales from 40,000 distinct stores belonging to 90 retail chains across 371 Metropolitan Statistical Areas (Argente et al., 2018). The 2010–2016 sample includes, on average, $234 billion in annual sales, representing roughly 2% of annual U.S. household consumption.

[7] More specifically, I summed sales each quarter for all unique combinations of brand_descr and product_module_descr in the Nielsen data. Results are robust to UPC-level analyses.

[8] The website is arguably among the most representative data sources for the résumés of white-collar workers in the United States. As of 2018, it had over 150 million U.S. profiles.

[9] The classification system uses profile information to classify individuals' job functions. Its classifications are not mutually exclusive (i.e., someone may be classified in both "product management" and "marketing"), which is important because different companies often assign different job titles to the same set of tasks (Baron & Bielby, 1985; Bechky & Chung, 2018; Sandholtz et al., 2019). My data-collection method imposed a cap of 1500 on the most relevant (according to the categorization algorithm) product- and research-related employees per firm; this restriction was only reached for one firm. In my sample, approximately 24% of the job titles of employees directly involved in the development and commercialization of products were explicitly labeled as research/R&D, 22% as product/innovation/strategy, 10% as marketing/brand, 9% as engineering, and the rest largely a mix of data analytics, IT, operations, and design. The networking website declined to provide further detail about its occupation-categorization algorithm; the company views the algorithm as a secret sauce for selling to HR professionals (its primary customers). The company claimed that the algorithm "has been refined over years and years of experimentation and research" by its teams of data scientists to make it as accurate as possible.

[10] There were more positions than unique employees because employees moved within and between firms.

source for measuring organizational culture,[11] there is precedent for using employee-generated text from online platforms, such as reviews of companies on Glassdoor (Bermiss & McDonald, 2018; Corritore et al., 2020; Marchetti, 2019; Moniz, 2015; Popadak, 2013).[12]

Using résumé text as a data source relies on a key assumption: that the language that individuals use to describe their work will accurately reflect the norms and values of their organizational cultures. This assumption is challenged by the possibility that people selectively compose aspirational job descriptions, rather than strictly factual accounts of their responsibilities. A later section will investigate this possibility and check the robustness of the measures. Here I will merely propose that the two most plausible reasons for mentioning, for example, "data analytics" in a résumé job description are (1) that the individual devoted considerable time and attention to data analytics in the position; or (2) that the individual perceived data analytics to be highly valued in the labor market and thus emphasized it within the limited space of the job description. Both reasons, aggregated over hundreds or thousands of employees, would tend to indicate an organizational culture that places a high normative value on quantitative analysis.

Finally, I obtained additional company information, used in controls and robustness checks, by obtaining all Glassdoor reviews for companies in my sample, and Compustat data for the public companies among them.

[11] Several studies have used résumé data to measure the adoption of quantitative tools and skills (Rock, 2021; Tambe, 2014; Wu, Hitt, et al., 2019; Wu, Lou, et al., 2019), but they have relied primarily on user-provided skills or job titles. By contrast, my measures draw on the text of employees' self-written job descriptions.

[12] Like the networking website I used, Glassdoor solicits free-text responses from employees to describe their work experiences. Whereas Glassdoor reviews may be appropriate for broad measures of the overall culture at a given company, job descriptions from résumés are a richer data source for measuring epistemic culture in particular, for two reasons. First, the topics in Glassdoor reviews (e.g., work-life balance, compensation) are less relevant to epistemic culture than are résumés—for instance, for companies in my sample, fewer than 1% of Glassdoor reviews mention data, whereas about 30% of résumés do. Second, the networking website offers a more comprehensive sample of employees than does Glassdoor, which allows me to compile a sufficient sample of employees who are directly related to product innovation.

To construct my primary dataset, I merged all of these data sources at the product level. (See Appendix A for additional details on sample construction). For additional analyses, I also constructed two datasets at different levels of analysis: a firm-level panel and an individual-level panel dataset. The product and firm-level datasets include all 3,502 new products launched by 61 consumer product manufacturers (NAICS codes 31, 32, and 33) in the Nielsen data for 2010–2016, retaining only firm-years with sufficiently large samples of employee résumés (over 100 product/research-related résumés). Because meaningful text analysis required using only firms with large numbers of innovation-related résumés, the final sample of firms represents only relatively large and established CPG firms. About 80% of the firms in the sample belonged to the Fortune 1000 or Global 500 in 2016.

Measures

Dependent Variable(s). The primary dependent variable, new-product sales, is operationalized as the sum of dollar sales of product *i* in the two quarters ending the 2-year post-launch period ($New\ Product\ Sales_i$). I used sales in the last two quarters of the 2-year post launch period to avoid confounding signals from the initial size of the product launch, which is more likely to be endogenously determined by the firm than later sales (Bass, 1969). In additional analyses, I discretize this measure to indicate whether the product was a commercial $Breakthrough_i$ (products in the top 5% of new-product sales in their product group) or a $Flop_i$ (products that earned less in the final two quarters of the 2-year post-launch period than in the initial 2 quarters). In the firm-year panel dataset, I sum these measures at the firm level for all products launched in each year.

Independent and Moderator Variables: Quantitative and Qualitative Analysis. To measure the extent to which each organizational culture values quantitative and qualitative analysis in the innovation process, I calculate the mean number of "qualitative" and "quantitative" words in employee résumés' innovation-related sentences. This measure draws on recent advances at the intersection of sociology and natural-language processing, which have established that a group's use of language reflects its cultural system (Kozlowski et al., 2019).

To construct this measure, I filtered the résumé text to include sentences with at least one word related to a product innovation (see Appendix B). Overall, 151,727 of 588,301 sentences (about 25%) in my résumé sample included a product-innovation-related word. The measures $\textit{Quantitative Analysis}_{ft}$ and $\textit{Qualitative Analysis}_{ft}$ were calculated for each firm f in year t by taking the mean number of quantitative and qualitative words within these product-innovation-related sentences.[13] I used a word embedding model to help identify relevant "quantitative" and "qualitative" words (Kozlowski et al., 2019; K. Li et al., 2021). For additional details on the methodologies I used to construct these measures, see Appendix B.

Table 1 displays the terms that most frequently occurred in the data, for both the quantitative and qualitative measures. Quantitative terms included words such as "data," "quantitative," "data analysis," and "data driven", and qualitative included words such as "qualitative," "interviews," "customer interaction," "user stories," "focus groups," and "ethnography."

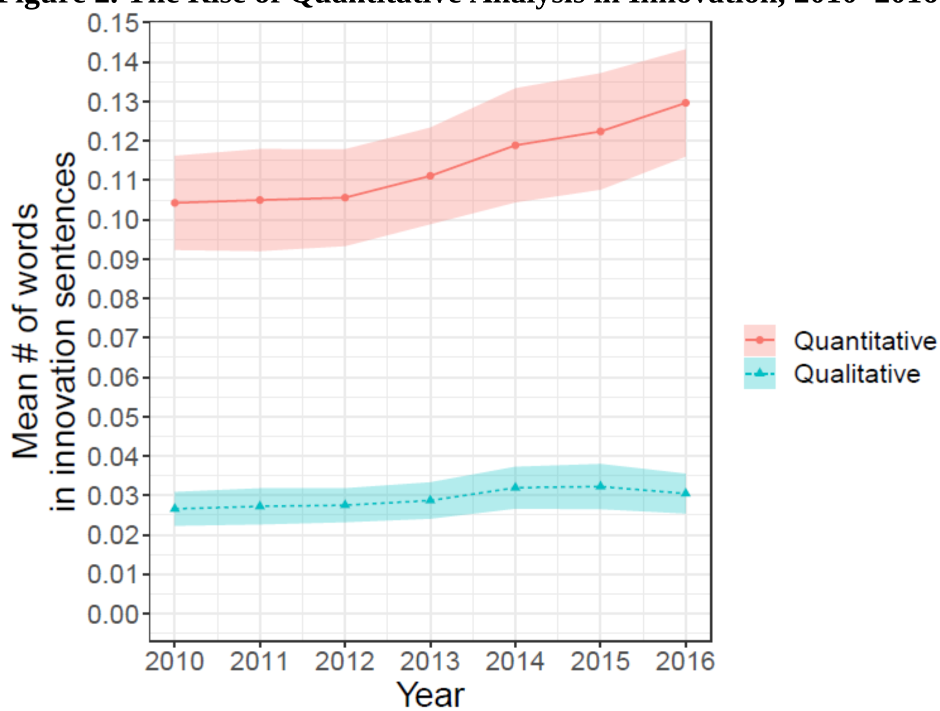
[13] As with other firm-level measures, I use a one-year lag: the time elapsed between a commitment to launch a given product and the actual launch is on average one year (6–18 months), a timeline confirmed in interviews with industry executives.

Table 1. Top 30 Quantitative and Qualitative terms by frequency

Quantitative	# of mentions	Qualitative	# of mentions
data	7083	*[interactions/visits/interview, etc.] within 5 words of [customer/consumer/field, etc.]*	1512
analytics	2039	qualitative	572
quantitative	592	user stories	463
survey	567	user acceptance test	333
bi	511	focus group	225
statistic	476	use case	195
dashboard	463	voice of customer	195
data analysis	376	voc	186
nielsen	366	uat	159
business intelligence	354	voice of the customer	132
market data	300	customer journey	101
sql	298	acceptance criteria	89
customer data	248	ethnograph	85
data driven	243	usability test	66
big data	226	personas	65
data analytics	193	case stud	58
consumer data	154	user story	55
tableau	151	face to face	43
sales data	133	persona	35
iri	120	consumer journey	33
nps	118	user journey	33
data science	102	epics	32
regress	95	storyboard	26
concept test	84	qual	24
gfk	68	one on one	22
consumer panel	67	epic	19
net promoter score	64	journey map	15
predictive analytic	60	field research	13
syndicated data	57	anthropolog	8
kantar	56	experience journey	8

Figure 2 validates these measures, displaying a steady increase in the mean use of quantitative analysis in innovation-related work between 2010 and 2016. This observation is consistent with the previous literature's observation of a quasi-exogenous increase in firms' adoption of quantitative tools and methods during this time period (Wu, Hitt, et al., 2019). I also validated the measures using an external source: mentions of data-related words in companies' earnings calls are significantly correlated with my quantitative analysis measure—both between companies and within companies over time (see Appendix C for analysis). Finally, I conducted an individual-level analysis demonstrating that individuals conformed to the organizational culture they joined: when an individual moves from an organization that uses less quantitative (or qualitative) analysis to one that uses more, his or her own individual use of that type of analysis significantly increases (See Appendix C for analysis). This pattern validates that the measure captures cultural fit and conformity (Bermiss & McDonald, 2018; Goldberg et al., 2016; Srivastava et al., 2018); it also helps validate my claim that my measure captures the culture around quantitative and qualitative analysis, as distinct from being merely the sum of individuals' skills.

Figure 2. The Rise of Quantitative Analysis in Innovation, 2010–2016



Alternative measure: Percent Quantitative. As an alternative to testing my hypothesis using separate independent and moderator variables, I calculate *Percent Quantitative*$_{ft}$ as the percentage of analysis that is quantitative: *Quantitative Analysis*$_{ft}$ / (*Quantitative Analysis*$_{ft}$ + *Qualitative Analysis*$_{ft}$). A moderate level of *Percent Quantitative* indicates balance between quantitative and qualitative (i.e., methodological pluralism); a high value indicates dominance of quantitative methods; and a low value indicates dominance of qualitative methods. This single variable composite of my independent and moderator variables is useful for representing the distribution, rather than levels, of quantitative and qualitative analysis. It is also useful for reducing the complexity of models with multiple interaction terms, and for the weighting procedure (described below).

Placebo measures. One potential concern is that the effects of quantitative and qualitative analysis may be confounded by a greater focus on the early-stage (i.e., inductive) vs. late-stage (i.e., deductive) portion of the innovation process. To address this possibility, I used the same process as for the quantitative and qualitative analysis variables to construct the placebo variables *Early Stage Analysis*$_{ft}$ and *Late Stage Analysis*$_{ft}$. (For additional details, see Appendix B.) In the robustness checks, I also check that the effects of quantitative and qualitative analysis (which I use as a proxy for organizational culture) are not fully explained by the quantitative and qualitative skills listed on employees' online profiles. I create placebo measures *Quantitative Skills*$_{ft}$ and *Qualitative Skills*$_{ft}$ by taking the mean proportion of employees' "skills" listed on their online profile that were related to quantitative and qualitative analysis.

Measures for Additional Analyses: Novelty, Market Uncertainty, and Intra-personal Methodological Pluralism. I created several measures to be used in additional analyses, which I describe later in the paper. I measure the novelty of product *i* as the percentage of Nielsen product attributes that were unique relative to all other products ever previously sold within the same product module (*Product Novelty*$_i$).[14] Due to incomplete product-attribute data, I was only able to calculate the novelty of a subset of products.[15] I calculate the market uncertainty of new product *i* as 1- R^2 from an OLS regression predicting sales in the pertinent Nielsen product group (*Market Uncertainty*$_i$).[16] Finally, for each firm-year I measure intra-personal methodological pluralism within employees as the mean of one minus the Herfindahl score of each individual's distribution of quantitative and qualitative words (*Intra-Personal Methodological Pluralism*$_{ft}$), following prior work on intra-personal cultural variation (Corritore et al., 2020; Marchetti, 2019). For example, someone who used 2 quantitative words and 1 qualitative word would have an intra-personal methodological pluralism of 1- $(2/3)^2$+$(1/3)^2 = 0.445$. This measure used only individuals who used at least one qualitative or quantitative word.

Control Variables. I control for several firm-year level characteristics, including the number of new products introduced by the firm that year, total annual sales (in $ millions), the total stock of products, and the number of profiles from the networking website. I lag each of these measures by one year, and take the natural log to normalize skewed distributions. Robustness

---

[14] In the Nielsen data, product modules are more granular product category classifications than product groups. Product novelty is more comparable within modules, because products within a module share the same number of potential product attributes. Consider two illustrative examples from the heavy-detergents product module, which has 6 relevant attributes: product type, form (e.g., liquid, liquid pac), container (e.g., bag, bottle), type (e.g., with bleach, with stain removal), scent, and size. When a new product enters the heavy-detergents category with a never-before-seen scent and container, its index is calculated as ("# new attributes") /("# detergent attributes") = (2)/(6) = 0.33. If the brand consists of multiple UPCs, the measure is calculated as the mean *Product Novelty*$_i$ of all UPCs launched in the first 2 years of the product's life. In additional analyses, I also measured product novelty using the hedonic pricing adjustments used in Argente, Lee and Moreira (2018) and Granja and Moreira (2019), with nearly identical results (correlation between measures >0.95).

[15] To ensure the quality of the measure, I used only products with high-quality product-attribute data. I included only product modules that had at least 50 distinct UPCs with over $100 in sales, and (after standardizing NA values and merging attribute values with at least 0.9 Jaro-Winkler textual similarity) kept only attribute values that appeared at least 5 times and attributes with at least 5 distinct values. I then kept only individual UPC codes that had non-NA values for at least 1/3 of their recorded product attributes, and had at least 3 recorded product attributes.

[16] I ran a separate OLS regression for each product group, using data at the product level. I regressed new product sales on the price and on the firm-level controls, with year fixed effects.

checks (presented in the Results section) include additional controls, such as "skills" and job titles, Glassdoor ratings, and public firm measures from Compustat such as capital and R&D expenditures.

Table 2 presents summary statistics for main variables used in the analysis; Table 3 presents a correlation matrix.

Table 2. Summary Statistics for Product-Level Data

	N	Mean	St. Dev.	Min	Pctl(25)	Median	Pctl(75)	Max
New-product sales (in $ millions)	3,502	1.395	4.365	0.000	0.007	0.226	1.127	98.407
Breakthrough	3,502	0.197	0.398	0	0	0	0	1
Flop	3,502	0.425	0.494	0	0	0	1	1
Quantitative analysis	3,502	0.110	0.040	0.028	0.085	0.107	0.121	0.353
Qualitative analysis	3,502	0.025	0.014	0.002	0.014	0.024	0.032	0.104
Percent quantitative	3,502	0.814	0.089	0.264	0.755	0.814	0.871	0.983
Product novelty	2,232	0.103	0.129	0.000	0.000	0.057	0.167	0.800
Market uncertainty	3,499	0.872	0.072	0.584	0.816	0.883	0.922	0.970
Intra-personal methodological pluralism	3,502	0.053	0.031	0.000	0.030	0.050	0.074	0.177
Profiles (in thousands)	3,502	1.061	0.703	0.154	0.424	0.999	1.402	2.603
Product introductions	3,502	9.818	9.375	0	3	7	14	51
Total sales (in $100 millions)	3,502	7.763	8.665	0	1.972	3.736	9.574	29.610
Total Products	3,502	416.119	282.566	0	171	402	636	960

Notes: Raw summary statistics at the product level. Logs and scaling are applied to some variables in regression analyses.

Table 3. Correlation Matrix for Product-Level Data

	1	2	3	4	5	6	7	8	9	10	11	12
1 New Product Sales ($M)	1											
2 Breakthrough	0.475	1										
3 Flop	-0.221	-0.315	1									
4 Quantitative analysis	0.053	0.085	0.019	1								
5 Qualitative analysis	-0.041	-0.020	0.022	0.090	1							
6 Percent quantitative	0.066	0.051	-0.014	0.359	-0.853	1						
7 Product novelty	-0.067	-0.061	0.097	0.016	0.108	-0.108	1					
8 Market uncertainty	-0.081	0.113	0.075	0.130	0.021	0.044	0.068	1				
9 Intra-personal methodological pluralism	0.003	-0.004	0.006	-0.065	0.577	-0.555	0.030	0.001	1			
10 Profiles (thousands)	0.101	0.001	-0.026	-0.072	0.132	-0.130	-0.088	-0.162	0.236	1		
11 Product introductions	-0.034	-0.099	0.031	-0.224	-0.047	-0.046	-0.063	-0.175	0.105	0.382	1	
12 Total sales ($100 M)	0.111	0.022	-0.004	0.071	0.214	-0.147	-0.037	-0.106	0.376	0.822	0.349	1
13 Total products	0.015	-0.069	-0.030	-0.148	-0.085	0.061	-0.147	-0.232	0.093	0.598	0.650	0.640

Statistical Estimation and Identification

An ideal experimental design would be to randomly assign varying degrees of quantitative analysis to organizations, and examine how it affects their innovation performance. Unfortunately, such a design was not feasible, nor was there a valid shock or instrumental variable available.

Though I do not claim strict causal identification, I attempt to mitigate potential endogeneity concerns throughout the paper. First, my models include an array of fixed-effects, including firm fixed-effects. Therefore, time-invariant inter-firm heterogeneity—such as a firm having a higher level of risk aversion or capital than others—will not bias the results. Second, I conduct an array of robustness checks with additional controls and specifications (described in a later section). Third, in robustness checks I attempt to further mitigate potential selection effects using Nonparametric Covariate Balancing Propensity Score (NPCBPS) weighting (Imai & Ratkovic, 2014). Unknown features of firms may predispose them simultaneously to be more data-driven and to produce fewer successful innovations. This concern is mitigated by weighting each observation by the inverse of its probability of treatment, as a function of confounding variables (Wang et al., 2021). Finally, I validate the theoretical underpinnings of my findings by conducting additional archival analyses (for example, examining product novelty and market uncertainty), and supplementing my quantitative results with qualitative support.

Throughout the paper, I use fixed-effects OLS models for continuous outcome variables (e.g., new product sales), conditional Logistic models for binary outcomes (e.g., breakthroughs and flops), and conditional fixed-effect Poisson models for count outcomes (e.g., counts of product introductions for firm-level analyses) (Wooldridge, 2010).

5. Results

Main Results

Table 4 reports coefficient estimates from regressions using product-level data with firm, product group, and year fixed-effects. Columns 1-4 report OLS regressions with *New Product Sales* as the dependent variable. Column 1 includes only *Quantitative Analysis* and control variables, indicating no statistically significant relationship between *Quantitative Analysis* and *New Product Sales*. Column 2 adds *Qualitative Analysis*, again indicating no statistically significant relationship.

I directly test my hypothesis in Table 4, column 3. I predicted that organizations' use of qualitative analysis would positively moderate the effect of quantitative analysis on the commercial success of new product innovations. Column 3 tests this hypothesis by adding the interaction term *Quantitative Analysis* × *Qualitative Analysis*. The positive and statistically significant coefficient on the interaction term ($p<0.01$) confirms the hypothesis, indicating that, on average, a new product launched by a firm with a one-standard-deviation increase in *Quantitative Analysis* is associated with a 38% ($e^{0.323}$ = 1.38) increase in sales for firms with *Qualitative Analysis* one standard deviation above the mean.

To ease interpretation of the coefficients, Figure 3 plots marginal effects from the models displayed in Table 4. Panel A displays marginal effects from the model in column 3: the predicted new product sales as a function of *Quantitative Analysis* on the x-axis from -1 (low) to 1 (high) standard deviations from the mean, for firms with *Qualitative Analysis* at -1 and 1 standard deviations from the mean. For products launched by companies with high levels of qualitative analysis (dotted blue line), increasing quantitative analysis from low to high increases predicted *New Product Sales* from about $90,000 to $200,000 (recall this represents sales in the final 2 quarters of a 2-year post-launch period). The opposite effect holds for low levels of qualitative analysis (solid red line): increasing quantitative analysis from low to high decreases predicted new product sales from about $150,000 to $100,000.

As an alternative test, Table 4, column 4 tests products' commercial success as a function of the percentage of the firm's analysis that was quantitative (*Percent Quantitative*). A moderate level of *Percent Quantitative* indicates balance between quantitative and qualitative (i.e., methodological pluralism); a high value indicates dominance of quantitative methods; and a low value indicates dominance of qualitative methods. Column 4 indicates a negative but not statistically significant relationship between *Percent Quantitative* and *New Product Sales*.

Table 4. Main Result; Product-level Analysis

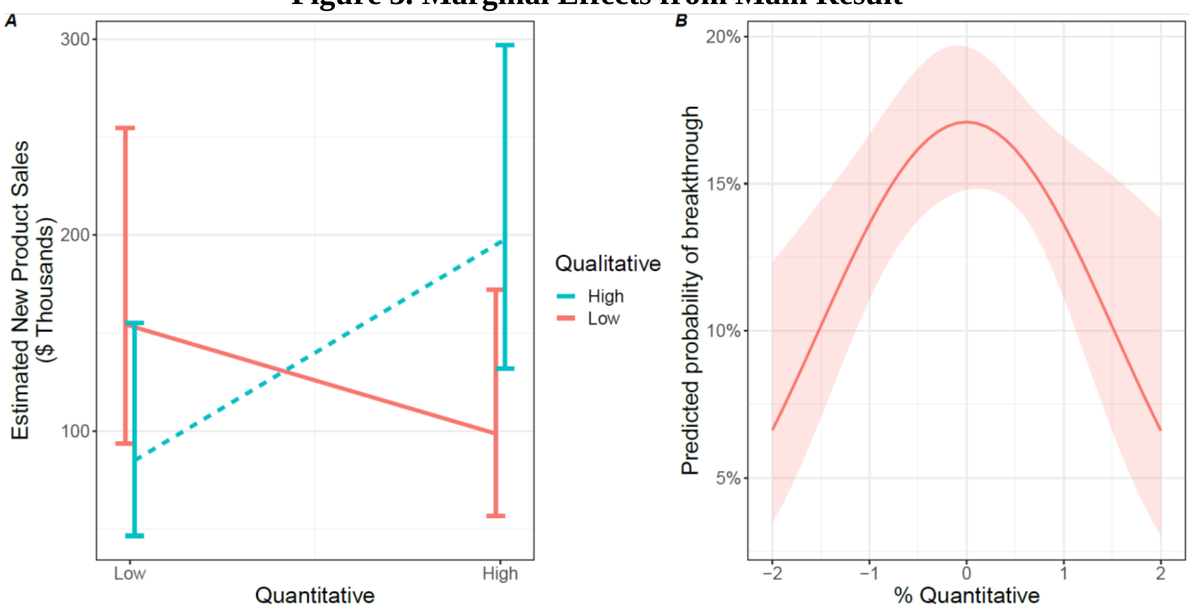
	OLS DV: New Product Sales				Logistic DV: Breakthrough	
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
Quantitative analysis	0.141	0.137	0.099		-0.035	
	(0.159)	(0.162)	(0.155)		(0.112)	
Qualitative analysis		0.043	0.024		-0.144	
		(0.181)	(0.184)		(0.146)	
Quant Analysis × Qual Analysis			**0.323****		**0.179***	
			(0.107)		**(0.079)**	
Percent Quantitative				-0.046		0.015
				(0.169)		(0.116)
Percent Quantitative Squared				**-0.154**		**-0.253***
				(0.099)		**(0.098)**
Total Analysis				0.167		-0.026
				(0.148)		(0.118)
Number of profiles	0.389	0.340	0.415	0.291	-0.748	-0.873*
	(0.538)	(0.611)	(0.661)	(0.603)	(0.470)	(0.430)
Product introductions	-0.453+	-0.454+	-0.451+	-0.445+	-0.366*	-0.355*
	(0.246)	(0.247)	(0.242)	(0.249)	(0.147)	(0.158)
Total sales	-0.068	-0.068	-0.062	-0.044	0.309	0.296
	(0.151)	(0.151)	(0.155)	(0.147)	(0.390)	(0.357)
Total products	0.338	0.323	0.551	0.425	0.430	0.415
	(0.749)	(0.744)	(0.802)	(0.738)	(0.674)	(0.658)
Company FE	YES	YES	YES	YES	YES	YES
Year FE	YES	YES	YES	YES	YES	YES
Product Group FE	YES	YES	YES	YES	YES	YES
Adj. R^2	0.117	0.116	0.118	0.117		
Within R^2	0.004	0.004	0.006	0.005		
Pseudo R^2					0.114	0.115
N	3502	3502	3502	3502	3326	3326

Notes: In columns 1-4, coefficients are estimated using OLS regression, with *New-Product Sales* as the dependent variable. In columns 5-6, coefficients are estimated using logistic regression, with *Breakthrough* as the dependent variable. The dependent variable (in OLS models) and control variables are log transformed to adjust for skew. The independent and moderator variables in interaction terms are z-score-scaled and mean-centered for ease of interpretation. Robust standard errors are clustered at the firm level. $^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$; $^{+}p < 0.1$.

According to my theory, methodological pluralism will especially benefit breakthrough innovations. Table 4, columns 5-6 report logistic regressions with *Breakthrough* as the dependent variable. Column 5 confirms that qualitative analysis positively moderates the effect of quantitative analysis on the likelihood of a breakthrough. Column 6 indicates a significant negative relationship for *Percent Quantitative Squared*, implying an inverted-U shaped relationship between *Percent Quantitative* and the likelihood of a breakthrough.

To ease the interpretation of the coefficients, Panel B in Figure 3 displays the predicted likelihood of a breakthrough as a function of *Percent Quantitative*. The figure confirms an inverted-U shape, indicating that breakthrough products are most likely to be launched by firms with moderate levels of *Percent Quantitative* (i.e., methodological pluralism).

Figure 3. Marginal Effects from Main Result



Notes: Marginal effects from models in Table 4. **Panel A:** Marginal effects of Quantitative×Qualitative interaction term on *New Product Sales* from Table 4 column 3. Low and High values represent 1 standard deviation below and above the mean respectively. The y-axis was rescaled by exponentiating the coefficient, then multiplying by 2 (Nielsen covers roughly 50% of U.S. grocery, drug. and mass retailers) and dividing by a thousand (to scale by thousands of dollars). **Panel B:** Marginal effects of *Percent Quantitative* on probability of *Breakthrough*, from Table 4 column 6.

Robustness Checks

I conducted several analyses to check the robustness of the main result. I report these analyses in Appendix D. First, I show that the main results are robust to using a discrete moderator variable (terciles of *Qualitative Analysis*) (Table D1 column 1), and to 1% Winsorization of all variables

(Table D1 columns 2 and 5). Second, I rule out potential concerns that the effects of quantitative and qualitative analysis are confounded by quantitative and qualitative "skills" rather than "culture," confounded by past years of quantitative experience or years of formal quantitative education, or confounded by early stage / inductive analysis (i.e., exploratory identification of new opportunities) vs. late stage / deductive analysis (i.e. testing well-defined options) (Table D1, columns 3 and 6). Third, I control for a broad array of variables derived from Compustat and Glassdoor: market value, ROA, net income, % of all Glassdoor job titles that were "R&D," % of all Glassdoor job titles that were "Data Analytics," Glassdoor mean overall rating, Glassdoor mean culture rating, earnings-call mentions of data in paragraphs about product or innovation, R&D expenditure, employment, and capital expenditures (Table D1, columns 4 and 7). Finally, I use the lagged, double lagged, and trend of each of these control variables to conduct Nonparametric Covariate Balancing Propensity Score (NPCBPS) to weight the independent variable *Percent Quantitative* (Table D1, column 8).[17] The results are robust to each of these additional analyses.

I also check that the results are robust to an alternative definition of the dependent variable *Breakthrough*. Throughout the paper, my main measure defines *Breakthrough* in terms of outlier commercial success. In Table D2, I use an alternative measure, *Novel Breakthrough*, defined as a product about the 95th percentile commercial success, which also had a *q/p* ratio above the median (Bass, 1969), and *Product Novelty* greater than 0. A high *q/p* ratio indicates a product that relied more on social diffusion than direct marketing, and is often indicative of relatively novel products that are not merely "replacement" products (Allen, 2022; Bass, 1969; Wang, 2019). Table D2

[17] I use the single variable *Percent Quantitative* (rather than the separate independent and moderator variables for quantitative and qualitative analysis) for the weighted model, because weights can only be assigned to one independent variable. A balance table is presented in Appendix D.

shows that the main results are driven by such *Novel Breakthroughs*, rather than *Other Breakthroughs* (i.e., breakthroughs with no novelty or a *q/p* ratio below the median).

Firm-Level Analysis

In addition to the product-level analysis presented above, I also present results from analyses on the firm-year panel dataset. The prior product-level analysis estimates the commercial success of individual products, conditional upon the product being launched; the firm-level analysis can capture a broader view—the overall innovation performance of firms over time. Table 5 displays the firm-level results.

Columns 1-3 test the effect of methodological pluralism using the *Quant Analysis* × *Qual Analysis* interaction term. Column 1 confirms a positive coefficient on *Quant Analysis* × *Qual Analysis* ($p<0.05$), indicating that for firms with *Qualitative Analysis* a standard deviation above the mean, a standard deviation increase in *Quantitative Analysis* is associated with a 93% increase in the sum of *New Product Sales* ($e^{0.66}$ = 1.93); Column 2 demonstrates that this increase in sales was not driven merely by launching a larger number of products. Column 3 confirms that firms with more methodological pluralism produced more breakthroughs (products in the 95th percentile of sales) ($p<0.05$); specifically, for firms with *Qualitative Analysis* a standard deviation above the mean, a standard deviation increase in *Quantitative Analysis* is associated with about a 20% increase in the number of breakthroughs.

Columns 4-6 test the effect of methodological pluralism using *Percent Quantitative*. Columns 4 and 5 show that a culture dominated by quantitative analysis is associated with fewer new products ($p<0.05$) and lower new product sales ($p<0.1$). Column 6 confirms a similar inverted-U pattern relationship between *Percent Quantitative* and *Breakthroughs* ($p<0.05$). Taken together, the firm-level results confirm that firms with more methodological pluralism produced a greater number of commercially successful new products, driven especially by breakthroughs.

Table 5. Firm Panel Results

	(1) OLS DV: New Product Sales	(2) Poisson DV: # New Products	(3) Poisson DV: # Breakthroughs	(4) OLS DV: New Product Sales	(5) Poisson DV: # New Products	(6) Poisson DV: # Breakthroughs
Quantitative analysis$_{ft-1}$	0.221	-0.085	0.020			
	(0.233)	(0.080)	(0.091)			
Qualitative analysis$_{ft-1}$	0.468*	0.173*	-0.108			
	(0.210)	(0.080)	(0.116)			
Quant Analysis$_{ft-1}$ × Qual Analysis$_{ft-1}$	0.666*	0.014	0.187*			
	(0.298)	(0.042)	(0.090)			
Percent Quantitative$_{ft-1}$				-0.591^{+}	-0.165*	0.131
				(0.299)	(0.070)	(0.122)
Percent Quantitative Squared$_{ft-1}$				-0.454	0.007	-0.271*
				(0.276)	(0.055)	(0.128)
Total Analysis$_{ft-1}$				0.478^{+}	0.011	0.005
				(0.255)	(0.071)	(0.114)
Company Controls	YES	YES	YES	YES	YES	YES
Company FE	YES	YES	YES	YES	YES	YES
Year FE	YES	YES	YES	YES	YES	YES
N	296	296	254	296	296	254
Adj. R^2	0.584			0.581		
Pseudo R^2		0.768	0.526		0.767	0.527

Notes: Coefficients are estimated using OLS regression (columns 1 and 4) and Poisson regression (columns 2,3,5, and 6) on firm-year panel (subscript *f* indicates firm-level variation and *t* indicates year-level variation). Lagged controls from Table 4 were summed to the firm level and included in the models: Number of profiles, Total Sales, and Total Products. The independent and moderator variables in interaction terms are z-score-scaled and mean-centered for ease of interpretation. Standard errors (in parentheses) are clustered at the firm level. $^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$; $^{+}p < 0.1$

Product Novelty as an Outcome and a Moderator

Next, I validate the theoretical underpinnings of my findings by demonstrating the role of product novelty, both as an outcome and as a moderating variable. My theory posits that it is relatively difficult to quantitatively capture the commercial potential of novel products ex ante. Therefore, I expect that more heavily quantitatively data-driven organizations will produce fewer novel innovations, and that attempts at producing novel innovations in such organizations will be less likely to succeed commercially.

Table 6 tests the role of novelty using the firm-year panel dataset. Columns 1-3 treat product novelty as an outcome variable. In Column 1, Percent Quantitative is negatively associated with the Mean Novelty of products launched by a firm, but the relationship is not statistically significant. Columns 2 and 3 reveal that Percent Quantitative is more negatively associated with highly novel products (those in the top quartile or top decile of novelty), as expected.

In Table 6, columns 4 and 5 treat novelty as a moderator of the relationship between Percent Quantitative and commercial success. In column 4, the coefficient on the interaction term Mean Novelty× Percent Quantitative suggests that a firm launching novel products will tend to produce fewer breakthrough commercial successes if they are highly quantitatively data-driven. Specifically, for a firm launching products with Mean Novelty a standard deviation above the mean, a standard deviation increase in Percent Quantitative is associated with about a 20% decrease in the number of breakthrough successes. Column 5 shows that there is no such significant interaction effect with flops as the outcome variable. But interestingly, a standard deviation increase in Percent Quantitative is associated with about a 30% decrease in flops. As might be expected given the inherent uncertainty in novel products, a standard deviation increase Mean Novelty is associated with about a 25% increase in flops.

Table 6. Product Novelty as an Outcome and Moderator

	Novelty as an Outcome			Novelty as a Moderator	
	(1) OLS DV: Mean Novelty	(2) Poisson DV: # 75th pctl Novel Products	(3) Poisson DV: # 90th pctl Novel Products	(4) Poisson DV: Breakthroughs	(5) Poisson DV: Flops
Percent Quantitative$_{ft-1}$	-0.224	-0.244*	-0.216^{+}	-0.086	-0.346**
	(0.134)	(0.105)	(0.120)	(0.109)	(0.107)
Percent Quantitative Squared$_{ft-1}$	-0.117	0.009	0.084	-0.056	0.142^{+}
	(0.084)	(0.093)	(0.099)	(0.115)	(0.084)
Mean Novelty$_{ft-1}$				0.029	0.226*
				(0.101)	(0.109)
Mean Novelty$_{ft-1}$× Percent Quantitative$_{ft-1}$				-0.232**	-0.103
				(0.085)	(0.108)
Mean Novelty$_{ft-1}$× Percent Quantitative Squared$_{ft-1}$				0.002	-0.092
				(0.087)	(0.085)
Total Analysis$_{ft-1}$	0.348*	0.038	0.095	0.039	-0.002
	(0.134)	(0.126)	(0.175)	(0.172)	(0.109)
Company Controls	YES	YES	YES	YES	YES
Company FE	YES	YES	YES	YES	YES
Year FE	YES	YES	YES	YES	YES
Adj. R^2	0.330				
Pseudo R^2		0.528	0.372	0.392	0.574
N	234	233	224	209	225

Notes: Coefficients are estimated using OLS (column 1) and Poisson regression (columns 2-5) on firm-year panel (subscript *f* indicates firm-level variation and *t* indicates year-level variation). Lagged controls from Table 4 were summed to the firm level and included in the models: Number of profiles, Total Sales, Total Intros, and Total Products. The independent and moderator variables in interaction terms are z-score-scaled and mean-centered for ease of interpretation. Standard errors (in parentheses) are clustered at the firm level. $^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$; $^{+}p < 0.1$

To briefly summarize: First, I observed that quantitative analysis is associated with a decreased rate of launching (especially highly novel) products. Second, I observed that organizations that relatively heavily use quantitative analysis (i.e., low methodological pluralism) produce fewer breakthrough successes when launching higher novelty products. Third, I observe that greater novelty is associated with an increased rate of flops, and dominant quantitative analysis with a decreased rate of flops. These observations are consistent the proposed role of novelty as an outcome and moderator in my theoretical argument.

Intra-personal Methodological Pluralism

Next, I explore how intra-personal methodological pluralism affects innovation performance. Consider two organizations which exhibit the same level of quantitative and qualitative analysis, in aggregate. The two organizations may differ in terms of how much quantitative and qualitative analysis varies within, rather than between, individuals. For example, the first organization may possess 50% of employees that use quantitative analysis, and 50% that use qualitative analysis; the second may possess 100% of employees that use both. To capture this distinction, I created the measure *Intra-Personal Methodological Pluralism*, which captures the degree to which individuals in the organization simultaneously use both methods, rather than being concentrated on using one or the other. I operationalized the measure by taking the mean of the inverse Herfindahl score of the distribution of mentions of quantitative and qualitative words for each individual, as described in the Methods section above (Corritore et al., 2020). Thus, the first organization (50% of employees use only each method) would have the lowest possible value of *Intra-Personal Methodological Pluralism*, while the second (all employees use both methods) would have the highest possible value.

In Table 7, I regress *Breakthrough* on *Intra-Personal Methodological Pluralism* using product-level data. Column 1 shows that on average, a standard deviation increase in *Intra-Personal Methodological Pluralism* is associated with about a 15% ($e^{-0.16}$=0.85) decrease in the probability of a breakthrough. But, column 2 adds the squared term, showing that the probability of a breakthrough has a significant inverted-U shaped relationship with *Intra-Personal Methodological Pluralism*. Column 3 shows that the inverted-U relationship holds even when adding the terms *Percent Quantitative* and *Percent Quantitative Squared* to the model. Column 4 confirms that there is also a significant inverted-U relationship between the number of breakthroughs and *Intra-Personal Methodological Pluralism* using a Poisson regression on the firm-year panel data.

Table 7. Intra-personal Methodological Pluralism

	Logistic; Product-level data; DV: Breakthrough			**Poisson; Firm panel data; DV: Breakthroughs**
	(1)	**(2)**	**(3)**	**(4)**
Intra-Personal Methodological Pluralism	-0.164**	-0.069	-0.122	-0.045
	(0.050)	(0.083)	(0.095)	(0.062)
Intra-Personal Methodological Pluralism Squared		-0.102***	-0.060**	-0.047*
		(0.018)	(0.020)	(0.019)
Total Analysis			-0.025	0.007
			(0.103)	(0.110)
Percent Quantitative			-0.122	
			(0.146)	
Percent Quantitative Squared			-0.230*	
			(0.105)	
Company FE	YES	YES	YES	YES
Product Group FE	YES	YES	YES	NA
Year FE	YES	YES	YES	YES
Pseudo R^2	0.114	0.115	0.117	0.566
N	3326	3326	3326	288

Notes: Coefficients in columns 1-3 were estimated using logistic regression on product-level data, with *Breakthrough* as the dependent variable. Column 4 was estimated using Poisson regression on firm-year panel data, with *Breakthroughs* as the dependent variable. Control variables are included, as in Table 4. Robust standard errors are clustered at the firm and year level (firm-level only for column 4). $^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$; $^{+}p < 0.1$

These results imply that a moderate amount of intra-personal pluralism helps innovation, but that too much may be detrimental. Although I cannot fully explain the underlying causes in this paper, one plausible explanation is that it is beneficial to specialize in a method, but it is also helpful to know enough of the other method to triangulate with its insights or productively collaborate with others who are using different methods in the organization. It is also important to note that, based on the results in Table 7, the intra-personal pluralism has an effect that is separate from the overall pluralism of the organization—a distinction I return to in the Discussion section.

Market Uncertainty as a Boundary Condition

According to my theory, I expect that organizations with low methodological pluralism will be less likely to produce breakthroughs under conditions of greater market uncertainty. This is because quantitative forecasts are much more reliable in products in relatively certain, stable markets (Allen, 2022; Felin & Zenger, 2017). Therefore, the degree of market uncertainty should be a boundary condition of my theory.

Table 8 tests this proposition on product-level data using a conditional logistic model that included the control variables and firm, year, and product group fixed-effects. In column 1, I first regressed *Breakthrough* on *Percent Quantitative* × *Market Uncertainty*. The result supports the intuition that products launched in markets with higher uncertainty are less likely to be breakthroughs if they are produced by firms with higher *Percent Quantitative*. In columns 2-5, I conduct subsample analyses using products launched in the 1st through 4th quartiles of product group market uncertainty. I regress *Breakthrough* on *Percent Quantitative* + *Percent Quantitative*^{2} to test when the main result holds. Only the 4th quartile of market uncertainty yields a significant negative coefficient for *Percent Quantitative*^{2}, supporting the proposed boundary condition that my theory is more applicable in conditions of higher levels of market uncertainty.

Table 8: Market Uncertainty as a Boundary Condition

DV: Breakthrough		Market Uncertainty Subsets (Quartiles)			
	(1) Full Dataset	**(2) 1st Quartile**	**(3) 2nd Quartile**	**(4) 3rd Quartile**	**(5) 4th Quartile**
Percent Quantitative	0.124	0.075	0.209	-0.137	-0.151
	(0.136)	(0.315)	(0.275)	(0.293)	(0.264)
Percent Quantitative×Market Uncertainty	-0.153^{*}				
	(0.077)				
Percent Quantitative Squared		-0.234	-0.530^{+}	-0.145	-0.175^{*}
		(0.263)	(0.303)	(0.415)	(0.082)
Total Analysis	-0.055	0.019	-0.328	0.323	0.018
	(0.100)	(0.250)	(0.229)	(0.349)	(0.138)
Controls	YES	YES	YES	YES	YES
Company FE	YES	YES	YES	YES	YES
Product Group FE	YES	YES	YES	YES	YES
Year FE	YES	YES	YES	YES	YES
Pseudo R^2	0.114	0.085	0.164	0.121	0.162
N	3324	791	760	816	815

Notes: Coefficients were estimated using logistic regression using product-level data, with *Breakthrough* as the dependent variable. Column 1 uses the whole dataset, and columns 2-5 use subsets of the data: products launched in the 1st-4th quartiles of product group uncertainty. Control variables are included, as in Table 4. The independent and moderator variables in interaction terms are z-score-scaled and mean-centered for ease of interpretation. Robust standard errors are clustered at the firm and year level. $^{**}p < 0.01$; $^{*}p < 0.05$; $^{+}p < 0.1$.

Antecedents of Data-Driven Culture

These results present a puzzle: if low-pluralism data-driven cultures are less successful at innovation, why are they so prevalent? This section presents a mixed-methods analysis suggesting that a rapid increase in quantitative analysis (without a commensurate increase in qualitative) was driven not by rational adoption, but by external management fads.

Industry informants traced the origins of their organizations' cultures to two interrelated sources: the external influence of management fads, and the internal influence of senior leadership. One manager, emphasizing the role of leadership, shared how difficult it was to present qualitative analysis to some leaders: "It's really hard to stand up in front of the senior management and they're like, 'Why do you think X?—And you don't have data, right?' . . . A lot of leaders call bullshit [on qualitative evidence] like, 'You got one guy telling you he doesn't like the color orange, and then you changed your mind?'" Employees took note of such leaders. One marketing manager remarked: "I know all the leaders in the building who don't really believe, who don't like focus groups, or whatever."

Other interviewees emphasized that the culture was shaped by external social forces. Many even displayed social-desirability bias during interviews. One interviewee exclaimed: "It's funny, when you asked me, 'Are you a data-driven company?' I was like, 'Well, yes, of course!' We're data-driven; that's what you say!" Combining both explanations, one senior innovation executive described how senior leaders mediated the influence of external fads:

> I think that there is this faith in big data right now, that it's hip and cool and fun to talk about at the country club, and CEOs say to themselves, 'Well, what are you doing?' 'Oh, I just started an analytics team.' . . . So I think some of it may also just be a trend. . . . It's just, it's cool.

Thus, leaders played a significant role in shaping which methods were deemed "acceptable" in each organization, but also appeared to be heavily influenced themselves by external fads that promoted quantitative decision-making.

This explanation appears to be supported by Figure 2, which shows a rapid rise in quantitative (but not qualitative) analysis between 2010 and 2016 that would otherwise be difficult to explain. One possible alternative explanation is that organizations rationally adopt more quantitative analysis via a process of organizational learning. However, this explanation was supported neither by the interview data nor by additional quantitative tests. Table 9 shows that, on the contrary, organizations may have learned from negative feedback to be less dependent on quantitative analysis (see Eggers, 2012; Eggers & Suh, 2018). Column 1 shows that, when firms with higher *Percent Quantitative* in time *t-2* produced more product flops in *t-1,* they became significantly less reliant on quantitative analysis in time *t*. But column 2 shows that, when firms with higher *Percent Quantitative* in time *t-2* produced more breakthrough successes in *t-1*, they did not differ statistically the following year. These patterns suggest that organizations did not becoming increasingly data-driven due to the technical merit or effectiveness of quantitative analysis; if anything, over time they learned to rebalance by relying less on quantitative analysis.

Table 9. Testing Organizational Learning as an Antecedent of Quantitative Culture

DV: Quantitative Analysis$_{ft}$	**(1)**	**(2)**
Flops$_{ft-1}$	-0.036	
	(0.078)	
Flops$_{ft-1}$ × Quantitative analysis$_{ft-2}$	-0.110*	
	(0.055)	
Breakthroughs$_{ft-1}$		-0.011
		(0.054)
Breakthroughs$_{ft-1}$ × Quantitative analysis$_{ft-2}$		-0.054
		(0.038)
Quantitative analysis$_{ft-2}$	0.132	0.136
	(0.100)	(0.109)
Company controls (double lag)	YES	YES
Company FE	YES	YES
Year FE	YES	YES
Adj. R^2	0.810	0.806
Within R^2	0.050	0.026
N	269	269

Notes: Coefficients are estimated using OLS regression on firm-year panel. Lagged control variables were included in the mdoels. Independent and moderator variables in interaction terms are z-score-scaled and mean-centered for ease of interpretation. Standard errors (in parentheses) are clustered at the firm level. $^{**}p < 0.01$; $^{*}p < 0.05$; $^{+}p < 0.1$

Indeed, without prompting, multiple informants at different companies independently used the word *pendulum* to describe the ebb and flow of data-driven culture. One manager related:

> From time to time—I want to say every 5–7 years—there's a bit of a counter-movement inside of our organization. And that is to basically pull us back from the extreme of being so data-driven, almost to the point of where it can become a barrier for us.

An aggregate picture emerges: external cultural trends increased firms' excessive reliance on quantitative analysis, but some organizations learned to rebalance in the direction of more methodologically pluralistic cultures. These observations suggest that the rise in data-driven organizational cultures may not be completely explainable by rational cost/benefit adoption, or by the emergence of a more sophisticated data-analytics ecosystem (Brynjolfsson & McElheran, 2016). Instead, my observations indicate that innovation in firms is embedded in a social context heavily influenced by external fads, triggers, and trends (Abrahamson & Fairchild, 1999; Bermiss et al., 2017; Zbaracki, 1998).

6. Discussion

This paper examines the impact of methodological pluralism on innovation in data-driven organizational cultures. Distinguishing between the *magnitude* of the use of quantitative methods and methodological *pluralism*, I hypothesized that qualitative analysis in the innovation process positively moderates the relationship between quantitative analysis and innovation success. Tests on a novel archival dataset of 3,502 new-product innovations launched by 61 CPG firms in 2010–2016 confirm my hypothesis. Challenging the intuition of prior literature, I found that the firms that used the most quantitative analysis also produced the most breakthrough innovations—but only if they also relied heavily on qualitative analysis. Supporting the theoretical underpinnings of my findings, I found quantitative analysis is associated with producing fewer highly novel products, and that organizations with low methodological pluralism produce fewer breakthrough successes when launching novel products. I also established high market uncertainty as a boundary condition, and provided evidence that external fads drove organizations to become excessively quantitative. These findings contribute new theoretical insight to three streams of research: organizational theories of innovation, theories of strategic decision-making, and research on the link between organizational culture and strategic performance.

Implications for Organizational Theories of Innovation: Transcending Exploration vs. Exploitation

In the literature on innovation in organizations, researchers have proposed a tradeoff: that data-driven organizations excel at commercializing incremental innovations (often associated with “exploitation”), but at the price of unintentionally deprioritizing less measurable breakthrough innovations (often associated with “exploration”) (Benner & Tushman, 2002; Can Deniz, 2020; Christensen & Bower, 1996). In contexts characterized by intra-organizational competition for scarce resources, they argue, organizations that use more quantitative analysis tend to allocate resources to relatively measurable—and thus relatively incremental—innovations.

I clarify that this tradeoff is only likely to be true when quantitative analysis is used to the exclusion of other methods. By distinguishing between the *magnitude* of an organization’s use of quantitative analysis and the methodological *pluralism* of its culture, I demonstrate that the actual innovation-stifling culprit is not the use of quantitative methods *per se*, but excessive focus on quantitative evidence at the expense of other (e.g., qualitative) evidence. In fact, in my analysis, the firms that were most successful at innovation were those that used the *most* quantitative analysis—but only if they also used qualitative analysis. This finding will come as a surprise to those familiar with decades of organizational research on exploration, exploitation, and the productivity dilemma (Abernathy, 1976; Baldwin & Clark, 1994; Benner & Tushman, 2003, 2015; Christensen & Bower, 1996) and with recent research on data analytics, experimentation, and innovation (Can Deniz, 2020; Felin et al., 2019; Felin & Zenger, 2017; Ghosh, 2021) which suggest that data-driven approaches are less likely to produce innovations that result in breakthrough commercial success.

Interviews with industry informants help to explain this surprising finding and to reconcile it with prior work. In keeping with prior innovation research, I did observe that a quantitatively data-driven culture could discourage breakthrough innovation. Some industry informants described a pervasive deference to “the data” that could discourage the questioning of assumptions and critical thinking necessary for identifying high-potential products. As one product manager observed, “The worst kind of decision making . . . is when you say that we're going to stick to our core revenue metrics for this evaluation; if it doesn't move those things, then we don't care.” A senior executive at another company shared a similar sentiment. In the past, she observed, her organization’s culture had highly valued and trusted the output of a handful of quantitative consumer tests: “It was kind of a ‘check the box.’ . . . People stopped digging deep into the richness of all the data we collected and just sort of checked the box and said, ‘We got a win on this one.’” In such a culture, it was difficult to mobilize resources to invest in potential breakthrough products.

But unlike prior innovation literature, my theory and findings do not imply a strict negative relationship between organizations’ use of quantitative data and breakthrough innovation. In particular, organizations characterized by methodological pluralism were able to use insights gleaned from quantitative data without necessarily accepting them at face value. Industry informants shared how pairing quantitative and qualitative analysis allowed for deeper critical thinking and questioning of assumptions. One manager observed: “The data can tell us what is happening, but the moment you start getting into the why it is happening, that's where the qualitative insights play a big role. . . . For us, it's the combination of the two which really gives us the breakthrough.” A manager at another company elaborated:

> Sometimes you get these little nuggets that you could never get in data, and you can use the qualitative to “see” your quantitative . . . just these little things that you would never really get, or you might ignore, because they'd be “standardized out” [in quantitative data]. Like when you're looking at the data, they don't pop. And then you can take those nuggets into either more quantitative or qualitative testing.

Triangulation of multiple methods was especially important for assessing potential breakthrough innovations. According to one R&D manager, when "inventing something totally new," the product would require "dealing with a consumer habit that we had to create and help her understand." Most feasible quantitative market tests don't capture "habit formation"—and would thus produce systematically more misleading findings about breakthroughs that require new habit formation. But rather than turning a blind eye to the quantitative, another R&D manager spoke of being "grounded in the qualitative"—interpreting all the information available rather than deferring to the quantitative test.

Thus, whereas prior research implied that innovation called for decreased use of quantitative measurement, my work suggests instead that innovation calls for more information and more analysis that varies in kind. One senior executive referred to this multi-method approach as a "body of evidence":

> So, before, the company was very unidirectional: did [the quantitative tests] give you good volume? Yes? Let's go—without even questioning whether the model is right, anything. . . . We started to look into how we can use the body of evidence. . . . You have to look at all the different signals, and connect all the data points, to confirm that this is a good idea . . . whether or not we would have some qualitative evidence [of excitement about the product] when you give it to consumers, and text analytics on the qualitative data, to see if they really love it or didn't. . . . So you use a lot of different small data points, but connected, that allow you to make a more informed decision.

In summary, my findings paint a new picture. By distinguishing between the use of quantitative analysis and methodological pluralism, my work challenges a longstanding conceptual intuition of the innovation literature: that organizations using quantitative analysis face a breakthrough/incremental tradeoff. I observed that—when viewed as one tool in a toolbox—more quantitative analysis can promote an organization's ability to commercialize breakthrough successes.

Implications for Theories of Strategic Decision-Making

My findings also contribute to research on strategic decision-making in organizations. Prior empirical studies in this literature almost universally demonstrate that data-driven models outperform humans in repeated-judgment contexts (Allen & Choudhury, 2022; Arthur & Hossein, 2019; Glaeser et al., 2021; Hoffman et al., 2018; Kesavan & Kushwaha, 2020; Kleinberg et al., 2018). Against this backdrop, I demonstrate an instructive anomaly: a repeated-judgment context (many large CPG firms produce dozens of new products per year) in which higher performance is characterized by methodological pluralism—arguably the opposite of compliance with a data-driven model.[18]

This observation advances our understanding of how compliance with data-driven models is dependent on context. According to my theory, pluralism (rather than compliance) leads to better innovation performance because quantitative forecasts are based on what worked in the past, or on snapshots of the present—but consumer and competitive dynamics can fundamentally change the "data-generating process" for what may drive breakthroughs in the future (Allen, 2022; Felin & Zenger, 2017, 2018; Zellweger & Zenger, 2021). Considering multiple different types of information causes innovators to question the implicit assumptions of the quantitative demand forecasts. The same skepticism toward market signals may in part explain why a scientific approach to innovation (i.e., hypothesis definition, validation, recognition of potential biases, etc.)

[18] The study did not directly measure compliance with data-driven models, but many of the most common quantitative analyses, such as concept tests and market-mix models, are highly algorithmic: a model translates a set of inputs into a score that determines whether a potential new product should be pursued. Furthermore, organizations' processes and routines can themselves be algorithmic (Csaszar & Eggers, 2013; Csaszar & Steinberger, 2020). Informants reported that in dogmatically data-driven organizations there was typically "one piece of data that is THE definitive answer"—a score on a concept test or market test that completely determined whether to invest in a new product. Investment decisions could be so dependent on such tests that, as one R&D manager recalled, "everyone was biting their nails [waiting for the test]."

in high uncertainty entrepreneurship is significantly more effective than a purely empirical or heuristic-driven approach (Camuffo et al., 2019; Felin et al., 2019).

An example from my field interviews will illustrate. Informants at one company in my sample told me that they had considered launching a new sanitizing laundry detergent, but the quantitative concept test had indicated low potential in the market. Incidentally, a competitor launched a highly similar product 6 months later, and successfully grew the category for several years. My informants' company repeated the same quantitative concept test five years later; this time it indicated that the product concept was an "outstanding" opportunity. The competitor's counterfactual success had demonstrated that the opportunity had in fact been quite large five years earlier; the test had indicated weak demand because the category didn't exist yet. Five years later, the opportunity appeared to be "outstanding" because it was more measurable—but in reality it would have become harder to launch a new product successfully, because the category was now highly competitive. Interpreting these empirical results correctly required theoretical understanding of competitive industry dynamics.

One possible objection to my argument is that, in light of recent advances in machine learning (ML) and artificial intelligence (AI), purely data-driven approaches could now use multiple types of evidence including unstructured qualitative data (Choudhury et al., 2021). Given all the same information available to humans (e.g., measuring facial reactions in a focus group), what is the difference between a data-driven model's use of multiple types of evidence and a human team's use of the same evidence? My theory and observations suggest that, even using multiple types of evidence, ML algorithms still have fundamental disadvantages relative to human innovators. For instance, although machine-learning (ML) algorithms can identify patterns in unstructured qualitative data, ML is not capable of an active orientation to data. It is purely empirical, and cannot use a scientific hypothesis-driven approach to interpret evidence in terms of an underlying theory (Camuffo et al., 2019). Nor can it question the data to learn from it, rather than merely reflecting or complying with it (Felin & Zenger, 2017; Kellogg et al., 2020; Leifer, 1992). It does not ask new questions, conduct new analyses as needed, apply learning across contexts, or draw connections between disparate forms of data in new ways—precisely the kinds of behaviors I qualitatively observed in organizations with high methodological pluralism. Humans using multiple forms of disparate evidence necessarily employ these skills, which machines are still far from achieving. In short, humans may outperform data-driven models in new-product innovation not only because they (may) have access to different types of information, but because of how flexibly they use and interpret the information in a dynamic environment. This line of reasoning raises intriguing possibilities for future research on both human and machine decision-making.

Implications for Organizational Culture and Strategic Performance

Finally, my study makes two contributions to the surprisingly sparse literature on the link between organizational culture and strategic performance (Chatman & O'Reilly, 2016; H. Li & van den Steen, 2019). First, though prior literature has examined how culture influences organizations' adaptability in changing environments (Chatman et al., 2014; Sørensen, 2002), the success of mergers and acquisitions (Bhatt, 2019; Marchetti, 2019), alignment of strategy with culture (Pisano, 1994, 2015), and foreign investments and market entries in global strategy (Kogut & Singh, 1988; Siegel et al., 2013), I present a novel way that culture impacts strategy: by shaping how organizational members perceive potential opportunities for innovation. This perspective has been repeatedly called for (Felin & Zenger, 2017; Furr & Eggers, 2021; Gavetti, 2012; H. Li & van den Steen, 2019; Puranam, 2018) but seldom examined empirically.

My findings demonstrate that, despite the outsize importance of breakthroughs in the CPG context—where producing breakthrough innovations is the explicit goal of top management—a surge in quantitatively data-driven organizational norms has driven investment that inadvertently led to less successful innovations (and fewer breakthroughs). This sequence of events suggests that consequential strategic decisions about which innovations to pursue are shaped as consequentially by organizations' cultures—specifically the norms that determine how potential opportunities are analyzed—as by deliberate top-down attempts at strategy.

Recognizing this link between culture and strategy allows for a fresh perspective on the role of strategic leaders in organizations. Whereas prior research has emphasized leaders' roles as decision-makers (Eisenhardt & Zbaracki, 1992), organizational designers (Csaszar & Eggers, 2013; Keum & See, 2017), and as external interfaces (Bermiss & Murmann, 2015; Khurana, 2011) in explaining firms' strategic performance, my study advances the view of leaders as orchestrating a cultural environment conducive to good strategic decision-making throughout the organization. I observed that top management influenced such decisions not by directly deciding which innovations to pursue, but by signaling which types of evidence—and thus which innovation opportunities—were valued most. For instance, one manager in my sample recounted that a chief executive viewed quantitative marketing-mix models as "the only true data to make a decision on whether to launch a product," which in turn encouraged organizational members to doggedly pursue projects that fit that description, with disastrous results. By contrast, other leaders actively designed a body-of-evidence approach that encouraged embracing the uncertainty inherent in using multiple kinds of analyses. In sum, my theory and observations expand the scope of strategy research by highlighting strategic leaders' indirect influence over innovation outcomes via the lever of their organizations' cultures.

Second, I contribute to the organizational culture literature by refining a conceptual dimension of organizational culture: the pluralism of content. Prior literature has focused on the *content* of specific norms, the *intensity* with which those norms were held to, and the *consensus* shared by organizational members (Chatman et al., 2014; Chatman & O'Reilly, 2016). For example, prior literature conceptualizes a world with two beliefs, A and B. The content dimension represents what A and B are (e.g., accountability and teamwork), intensity represents how strongly beliefs about A and B are held, and inter- or intra-personal consensus represents the distribution of these beliefs among members (Corritore et al., 2020). My paper further refines the *content* dimension by demonstrating that the combination of specific norms matters—having A and B together may be different than their individual effects. This combination of values may be especially important when A and B represent polarizing beliefs, and do not easily coexist (e.g., progressive and conservative values; quantitative and qualitative methods) except in a pluralistic culture. Such pluralism appears to be the driving force of the relationship between data-driven organizational culture and innovation, but has not been emphasized in existing conceptualizations of organizational culture.

Limitations and Future Directions

This study's limitations present promising avenues for future research. First, the data sample was necessarily restricted to large CPG firms; the theory may not generalize to other industries (e.g., B2B software) or to less-mature organizations (e.g., startups) in relatively information-poor environments (Hallen & Pahnke, 2016; Huang & Pearce, 2015). It may be particularly useful for future research to explore how cultures form around qualitative and quantitative evidence in the relatively volatile setting of high-growth startups. Because startups are more accustomed to uncertainty (Bhide, 2003), they may suffer less from hindrances to effective resource allocation than do large organizations (Burgelman, 1991; Choudhury, 2017; Christensen & Bower, 1996; Pisano, 2019). In light of my theory and observations, I expect startups to face the same cognitive constraints (i.e., difficulty perceiving breakthrough opportunities, neglecting to question the assumptions behind quantitative evidence) but fewer organizational constraints (i.e., difficulty obtaining internal resources for less quantitatively measurable initiatives). These dimensions could be isolated, and this proposition tested, in future research.

A second limitation of this study is its inability to fully address all potential concerns about causal identification, or to directly measure all the mechanisms at play. I attempted to address as many endogeneity concerns as possible, for example by using an array of fixed effects, using weighting to eliminate observable treatment-selection heterogeneity, and validating theoretical underpinnings in additional analyses. However, the lack of a causal intervention leaves the possibility of unobserved omitted variables. Furthermore, I was not able to observe all projects that firms initiated, or to measure on an individual or group level whether the proposed cognitive mechanisms (i.e., difficulty perceiving breakthrough opportunities when neglecting to question assumptions behind quantitative evidence) or organizational mechanisms (i.e., difficulty obtaining internal resources for less quantitatively measurable initiatives) drove the results. This provides opportunities for future research.

7. Conclusion

The prevalence and influence of data-driven innovation has increased dramatically in recent years. Prior work, focused on the magnitude of organizations' use of data, has suggested a tradeoff: that more data-driven organizations are more effective at incremental innovation but less likely to produce breakthroughs. By recognizing the possibility of methodological pluralism—organizations that use more quantitative analysis may also use more qualitative analysis—I show that this tradeoff can be transcended. Organizations whose cultures highly value quantitative analysis of market demand are most likely to produce breakthrough commercial successes, but only if they are pluralistic enough to also value qualitative analysis. Although an organizational culture's reliance on qualitative and quantitative analysis is heavily influenced by external cultural trends, leaders can shape their culture as an indirect lever for perceiving and capturing innovation opportunities.

References

Abernathy, W. J. (1976). *The Productivity Dilemma: Roadblock to Innovation in the Automobile Industry*.

Abrahamson, E., & Fairchild, G. (1999). Management fashion: Lifecycles, triggers, and collective learning processes. *Administrative Science Quarterly*, *44*(4), 708–740.

Ahuja, G., & Lampert, C. M. (2001). Entrepreneurship in the large corporation: A longitudinal study of how established firms create breakthrough inventions. *Strategic Management Journal*, *22*(6–7), 521–543.

Akerlof, G. A. (2020). Sins of Omission and the Practice of Economics. *Journal of Economic Literature*, *58*(2), 405–418.

Allen, R. T. (2022). The Market-Size Paradox of Niche Market Innovations. In *Academy of Management Proceedings* (Vol. 2022, Issue 1). Academy of Management Briarcliff Manor, NY 10510.

Allen, R. T., & Choudhury, P. (2022). Algorithm-Augmented Work and Domain Experience: The Countervailing Forces of Ability and Aversion. *Organization Science*, *December*.

Anthony, C. (2018). To question or accept? How status differences influence responses to new epistemic technologies in knowledge work. *Academy of Management Review*, *43*(4), 661–679.

Anthony, C. (2021). When Knowledge Work and Analytical Technologies Collide: The Practices and Consequences of Black Boxing Algorithmic Technologies. *Administrative Science Quarterly*, 000183922110167.

Argente, D., Lee, M., & Moreira, S. (2018). Innovation and product reallocation in the great recession. *Journal of Monetary Economics*, *93*, 1–20.

Arthur, F., & Hossein, K. R. (2019). Deep learning in medical image analysis: A third eye for doctors. *Journal of Stomatology, Oral and Maxillofacial Surgery*.

Baldwin, C., & Clark, K. B. (1994). Capital-Budgeting Systems and Capabilities Investments in U.S. Companies after the Second World War. *Business History Review, 68*(1), 1960–1990.

Baron, J. N., & Bielby, W. (1985). Organizational barriers to gender equality. *Gender and the Life Course*.

Bass, F. M. (1969). A new product growth for model consumer durables. *Management Science*, *15*(5), 215–227.

Bechky, B. A., & Chung, D. E. (2018). Latitude or latent control? How occupational embeddedness and control shape emergent coordination. *Administrative Science Quarterly*, *63*(3), 607–636.

Benner, M. J., & Tushman, M. (2002). Process Management and Technological Innovation: A Longitudinal Study of the Photography and Paint Industries. *Administrative Science Quarterly*, *47*(4), 676.

Benner, M. J., & Tushman, M. L. (2003). Exploitation , Exploration , and Process Management: The Productivity Dilemma Revisited. *The Academy of Management Review*, *28*(2), 238–256.

Benner, M. J., & Tushman, M. L. (2015). Reflections on the 2013 Decade Award - “Exploitation, Exploration, and Process Management: The Productivity Dilemma Revisited” ten years later. *Academy of Management Review*, *40*(4), 497–514.

Berg, J. M., & Yu, A. (2021). Getting the picture too late: Handoffs and the effectiveness of idea implementation in creative work. *Academy of Management Journal*, *64*(4), 1191–1212.

Bermiss, S., Hallen, B. L., McDonald, R., & Pahnke, E. C. (2017). Entrepreneurial beacons: The Y ale endowment, run-ups, and the growth of venture capital. *Strategic Management Journal*, *38*(3), 545–565.

Bermiss, S., & McDonald, R. M. (2018). Ideological misfit? Political affiliation and employee departure in the private-equity industry. *Academy of Management Journal*, *61*(6), 2182–2209.

Bermiss, S., & Murmann, J. P. (2015). Who matters more? The impact of functional background and top executive mobility on firm survival. *Strategic Management Journal*, *36*(11), 1697–1716.

Beunza, D., & Stark, D. (2004). Tools of the Trade: The Socio-Technology of Arbitrage in a Wall Street Trading Room. *Industrial and Corporate Change*, *13*(2), 253–290.

Bhatt, A. (2019). *Culture code-switching in a post-merger organization*.

Bhide, A. (2003). *The origin and evolution of new businesses*. Oxford University Press.

Bower, J. (1970). *Managing the Resource Allocation Process: A Study of Corporate Planning and Investment*.

Brynjolfsson, E., Hitt, L. M., & Kim, H. H. (2011). Strength in Numbers: How Does Data-Driven Decisionmaking Affect Firm Performance? *Ssrn*.

Brynjolfsson, E., Jin, W., & McElheran, K. S. (2021). The Power of Prediction: Predictive Analytics, Workplace Complements, and Business Performance. *SSRN Electronic Journal*, 1–42.

Brynjolfsson, E., & Mcelheran, K. (2019). Data in Action: Data-Driven Decision Makingand Predictive Analytics in U.S. Manufacturing. *SSRN Electronic Journal*, 1–55.

Brynjolfsson, E., & McElheran, K. (2016). The Rapid Adoption of Data-Driven Decision-Making. *American Economic Review*, *106*(5), 133–139.

Burgelman, R. A. (1991). Intraorganizational Ecology of Strategy Making and Organizational Adaptation: Theory and Field Research. *Organization Science*, *2*(3), 239–262.

Camuffo, A., Cordova, A., Gambardella, A., & Spina, C. (2019). A Scientific Approach to Entrepreneurial Decision Making: Evidence from a Randomized Control Trial.

Management Science, *October*.
Can Deniz, B. (2020). Experimentation and Incrementalism: The Impact of the Adoption of A/B Testing. *National Bureau of Economic Research*.
Chatman, J. A., Caldwell, D. F., O'Reilly, C. A., & Doerr, B. (2014). Parsing organizational culture: How the norm for adaptability influences the relationship between culture consensus and financial performance in high-technology firms. *Journal of Organizational Behavior, 35*(6), 785–808.
Chatman, J. A., & O'Reilly, C. A. (2016). Paradigm lost: Reinvigorating the study of organizational culture. *Research in Organizational Behavior*, *36*, 199–224.
Choudhury, P. (2017). Innovation outcomes in a distributed organization: Intrafirm mobility and access to resources. *Organization Science*, *28*(2), 339–354.
Choudhury, P., Allen, R. T., & Endres, M. G. (2021). Machine Learning for Pattern Discovery in Management Research. *Strategic Management Journal*.
Choudhury, P., & Haas, M. R. (2018). Scope versus speed: Team diversity, leader experience, and patenting outcomes for firms. *Strategic Management Journal*, *39*(4), 977–1002.
Christensen, C. M., & Bower, J. L. (1996). Customer Power, Strategic Investment, and the Failure of Leading Firms. *Strategic Management Journal*, *17*(3), 197–218.
Conti, R., Godinho de Matos, M., & valentini, giovanni. (2020). Big for Everyone? Big Data, Firm Size and Performance. *SSRN Electronic Journal*, 1–45.
Corritore, M., Goldberg, A., & Srivastava, S. B. (2020). Duality in Diversity: How Intrapersonal and Interpersonal Cultural Heterogeneity Relate to Firm Performance. *Administrative Science Quarterly*, *65*(2), 359–394.
Csaszar, F. A., & Eggers, J. P. (2013). Organizational decision making: An information aggregation view. *Management Science*, *59*(10), 2257–2277.
Csaszar, F. A., & Steinberger, T. (2020). Organizations as Artificial Intelligences: The Use of Artificial Intelligence Analogies in Organization Theory. *Academy of Management Annals*.
Denrell, J., Fang, C., & Winter, S. G. (2003). The economics of strategic opportunity. *Strategic Management Journal*, *24*(10 SPEC ISS.), 977–990.
DeSantola, A., Gulati, R., & Zhelyazkov, P. I. (2020). *External Interfaces and Internal Processes: Market Positioning and Divergent Professionalization Paths in Young Ventures*.
Edmondson, A., & Mcmanus, S. E. (2007). Methodological fit in management. *Academy of Management Review*, *32*(4), 1155–1179.
Eggers, J. P. (2012). All experience is not created equal: Learning, adapting, and focusing in product portfolio management. *Strategic Management Journal*, *33*(3), 315–335.
Eggers, J. P., & Suh, J.-H. (2018). Experience and Behavior: How Negative Feedback in New Versus Experienced Domains Affects Firm Action and Subsequent Performance. *Academy of Management Journal*, *62*(2), 309–334.
Eisenhardt, K. M., & Zbaracki, M. J. (1992). Strategic Decision Making. *Strategic Management Journal*, *13*, 17–37.
Feldman, M. S., & March, J. G. (1981). Information in Organizations as Signal and Symbol. *Administrative Science Quarterly*, *26*(2), 171–186.
Felin, T., Gambardella, A., Stern, S., & Zenger, T. (2019). Lean startup and the business model: Experimentation revisited. *Long Range Planning*, *March*, 101889.
Felin, T., & Zenger, T. (2017). The Theory-Based View: Economic Actors as Theorists. *Strategy Science*, *2*(4), 258–271.
Felin, T., & Zenger, T. (2018). What sets breakthrough strategies apart. *MIT Sloan Management*

Review, *59*(2), 86–88.
Furr, N. R., & Eggers, J. P. (2021). Behavioral innovation and corporate renewal. *Strategic Management Review*, *2*(2), 285–322.
Gao, C., & McDonald, R. (2022). Shaping nascent industries: Innovation strategy and regulatory uncertainty in personal genomics. *Administrative Science Quarterly*.
Gavetti. (2012). Toward a behavioral theory of strategy. *Organization Science*, *23*(1), 267–285.
Gavetti, G., Helfat, C. E., & Marengo, L. (2017). Searching, shaping, and the quest for superior performance. *Strategy Science*, *2*(3), 194–209.
Gavetti, G., & Levinthal, D. (2000). Looking forward and looking backward: Cognitive and experiential search. *Administrative Science Quarterly*, *45*(1), 113–137.
Ghosh, S. (2021). *Think Before You Act: The Unintended Consequences of Inexpensive Experimentation*.
Gilbert, C. G., & Bower, J. L. (2005). *From resource allocation to strategy*. Oxford University Press.
Glaeser, E. L., Hillis, A., Kim, H., Kominers, S. D., & Luca, M. (2021). *Decision Authority and the Returns to Algorithms*.
Goldberg, A., Srivastava, S. B., Manian, V. G., Monroe, W., & Potts, C. (2016). Fitting in or standing out? The tradeoffs of structural and cultural embeddedness. *American Sociological Review*, *81*(6), 1190–1222.
Granja, J., & Moreira, S. (2019). Product Innovation and Credit Market Disruptions. *SSRN Electronic Journal*.
Hallen, B. L., & Pahnke, E. C. (2016). When do entrepreneurs accurately evaluate venture capital firms' track records? A bounded rationality perspective. *Academy of Management Journal*, *59*(5), 1535–1560.
Hitt, L., Jin, F., & Wu, L. (2015). Data Skills and Value of Social Media: Evidence from Large-Sample Firm Value Analysis. *Working Paper*, 1–19.
Hoffman, M., Kahn, L. B., & Li, D. (2018). Discretion in hiring. *The Quarterly Journal of Economics*, *133*(2), 765–800.
Huang, L., & Pearce, J. L. (2015). Managing the Unknowable: The Effectiveness of Early-stage Investor Gut Feel in Entrepreneurial Investment Decisions. In *Administrative Science Quarterly* (Vol. 60, Issue 4).
Imai, K., & Ratkovic, M. (2014). Covariate balancing propensity score. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *76*(1), 243–263.
Jick, T. D. (1979). Mixing Qualitative and Quantitative Methods: Triangulation in Action. *Administrative Science Quarterly*, *24*(4), 602–611.
Kaplan, S. (2011). Strategy and Powerpoint: An Inquiry into the Epistemic Culture and Machinery of Strategy Making. *Organization Science*, *22*(2), 320–346.
Kaplan, S. (2016). Mixing quantitative and qualitative research. *Handbook of Qualitative Organizational Research: Innovative Pathways and Methods*, 423–433.
Kellogg, K. C., Valentine, M. A., & Christin, A. (2020). Algorithms at work: The new contested terrain of control. *Academy of Management Annals*, *14*(1), 366–410.
Kesavan, S., & Kushwaha, T. (2020). Field Experiment on the Profit Implications of Merchants' Discretionary Power to Override Data- Driven Decision-Making Tools. *Management Science*.
Keum, D., & See, K. E. (2017). The Influence of Hierarchy on Idea Generation and Selection in the Innovation Process. *Organization Science*, *28*(4), 653–669.

Khurana, R. (2011). *Searching for a corporate savior*. Princeton University Press.
Kleinberg, J., Lakkaraju, H., Leskovec, J., Ludwig, J., & Mullainathan, S. (2018). Human decisions and machine predictions. *The Quarterly Journal of Economics*, *January*, 237–293.
Klingebiel, R., & Rammer, C. (2014). Resource allocation strategy for innovation portfolio management. *Strategic Management Journal*, *35*(2), 246–268.
Knight, F. H. (1921). *Risk, uncertainty and profit* (Vol. 31). Houghton Mifflin.
Knorr-Cetina, K. (1999). *Epistemic Cultures: How the Sciences Make Knowledge*.
Kogut, B., & Singh, H. (1988). The effect of national culture on the choice of entry mode. *Journal of International Business Studies*, *19*(3), 411–432.
Koning, R., Hasan, S., & Chatterji, A. (2022). Experimentation and Startup Performance: Evidence from A/B Testing. *Management Science*.
Kozlowski, A. C., Taddy, M., & Evans, J. A. (2019). The Geometry of Culture: Analyzing the Meanings of Class through Word Embeddings. *American Sociological Review*, *84*(5), 905–949.
Lamont, M. (2009). *How professors think*. Harvard University Press.
Leifer, E. M. (1992). Denying the data: Learning from the accomplished sciences. *Sociological Forum*, *7*(2), 283–299.
Leonard-Barton, D. (1992). Core capabilities and core rigidities: a paradox in managing new product development. *Strategic Management Journal*, *13*(Strategy Process: Managing Corporate Self-Renewal (Summer, 1992)), 111–125.
Levinthal, D. A. (1997). Adaptation on Rugged Landscapes. *Management Science*, *43*(7), 934–950.
Levinthal, D. A. (2007). Bringing selection back into our evolutionary theories of innovation. *Perspectives on Innovation*, *February*, 293–307.
Levinthal, D. A., & March, J. G. (1993). The myopia of learning. *Strategic Management Journal*, *14*(2 S), 95–112.
Li, H., & van den Steen, E. (2019). Birds of a Feather . . . Enforce Social Norms? Interactions Among Culture, Norms, and Strategy. *SSRN Electronic Journal*.
Li, K., Mai, F., Shen, R., & Yan, X. (2021). Measuring Corporate Culture Using Machine Learning. *Review of Financial Studies*, *34*(7), 3265–3315.
Lix, K., Goldberg, A., Srivastava, S., & Valentine, M. (2019). *Expressly Different: Discursive Diversity and Team Performance*.
March, J. (1991). Exploration and Exploitation in Organizational Learning. *Organization Science*, *2*(1), 71–87.
Marchetti, A. (2019). *Firms of a Feather Merge Together: The Role of Acquirer-Target Culture Compatibility in Technology Acquisitions*. https://www.dropbox.com/s/8xzxqch29fxsnwl/JobMarketPaper_AriannaMarchetti.pdf?dl=0
Martin, R. L., & Golsby-Smith, T. (2017). Management Is Much More Than a Science. *Harvard Business Review*, *October 2017*.
Mazmanian, M., & Beckman, C. M. (2018). "Making" your numbers: Engendering organizational control through a ritual of quantification. *Organization Science*, *29*(3), 357–379.
McAfee, A., & Brynjolfsson, E. (2012). Big Data: The Management Revolution. *Harvard Business Review*, *October*. http://tarjomefa.com/wp-content/uploads/2017/04/6539-English-TarjomeFa-1.pdf
McDonald, R. M., & Allen, R. T. (2021). A Spanner in the Works: Category-Spanning Entrants

and Audience Valuation of Incumbents. *Strategy Science*, *May*.
McDonald, R. M., & Eisenhardt, K. M. (2019). Parallel Play: Startups, Nascent Markets, and Effective Business-model Design. *Administrative Science Quarterly*, 000183921985234.
McDonald, R. M., & Gao, C. (2019). Pivoting Isn't Enough? Managing Strategic Reorientation in New Ventures. *Organization Science, November*.
Moniz, A. (2015). Inferring employees' social media perceptions of goal-setting corporate cultures and the link to firm value. *Unpublished Working Paper*, *October*.
Müller, O., Fay, M., & vom Brocke, J. (2018). The Effect of Big Data and Analytics on Firm Performance: An Econometric Analysis Considering Industry Characteristics. *Journal of Management Information Systems*, *35*(2), 488–509.
O'Reilly, C., & Tushman, M. L. (2008). Ambidexterity as a dynamic capability: Resolving the innovator's dilemma. *Research in Organizational Behavior*. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=978493
Orlikowski, W. J. (1992). Learning from Notes: Organizational Issues in Groupware Implementation. *Proceedings of the 1992 ACM Conference on Computer-Supported Cooperative Work*.
Orlikowski, W. J. (2007). Sociomaterial practices: Exploring technology at work. *Organization Studies*, *28*(9), 1435–1448.
Page, S. E. (2018). Why "Many-Model Thinkers" Make Better Decisions. *Harvard Business Review*.
Pahnke, E. C., McDonald, R. M., Wang, D., & Hallen, B. (2015). Exposed: Venture capital, competitor ties, and entrepreneurial innovation. *Academy of Management Journal*, *58*(5), 1334–1360.
Patel, N. (2013). *How Netflix Uses Analytics To Select Movies, Create Content, & Make Multimillion Dollar Decisions*. Neilpatel.Com. https://blog.kissmetrics.com/how-netflix-uses-analytics/
Pisano, G. P. (1994). Knowledge, integration, and the locus of learning: An empirical analysis of process development. *Strategic Management Journal*, *15*(S1), 85–100.
Pisano, G. P. (2015). You need an innovation strategy. *Harvard Business Review*, *93*(6), 44–54.
Pisano, G. P. (2019). *Creative construction: The DNA of sustained innovation*. PublicAffairs.
Popadak, J. A. (2013). A Corporate Culture Channel: How Increased Shareholder Governance Reduces Firm Value. *SSRN Electronic Journal*, *19104*.
Puranam, P. (2018). The Organizational Foundations of Behavioral Strategy. In *Behavioral Strategy in Perspective* (Vol. 39, pp. 79–89). Emerald Publishing Limited.
Rindova, V., & Courtney, H. (2020). To shape or adapt: Knowledge problems, epistemologies, and strategic postures under knightian uncertainty. *Academy of Management Review*, *45*(4), 787–807.
Rock, D. (2021). Engineering Value: The Returns to Technological Talent and Investments in Artificial Intelligence. *SSRN Electronic Journal*, 1–85.
Sandholtz, K., Chung, D., & Waisberg, I. (2019). The double-edged sword of jurisdictional entrenchment: explaining human resources professionals' failed strategic repositioning. *Organization Science*, *30*(6), 1349–1367.
Schein, E. H. (2010). *Organizational culture and leadership* (Vol. 2). John Wiley & Sons.
Schoenle, R. (2017). International menu costs and price dynamics. *Review of International Economics*, *25*(3), 578–606.
Seemann, J. (2012). Hybrid insights: Where the quantitative meets the qualitative. *Rotman*

Magazine, *57–61*.
Shapiro, J. (2018). Help Your Team Understand What Data Is and Isn't Good For. *Harvard Business Review*.
Siegel, J. I., Licht, A. N., & Schwartz, S. H. (2013). Egalitarianism, cultural distance, and foreign direct investment: A new approach. *Organization Science*, *24*(4), 1174–1194.
Siggelkow, N. (2007). Persuasion with Case Studies. *Academy of Management Journal*, *50*(1), 20–24.
Singh, J., & Fleming, L. (2010). Lone inventors as sources of breakthroughs: Myth or reality? *Management Science*, *56*(1), 41–56.
Sørensen, J. B. (2002). The strength of corporate culture and the reliability of firm performance. *Administrative Science Quarterly*, *47*(1), 70–91.
Srivastava, S. B., Goldberg, A., Manian, V. G., & Potts, C. (2018). Enculturation trajectories: Language, cultural adaptation, and individual outcomes in organizations. *Management Science*, *64*(3), 1348–1364.
Tambe, P. (2014). Big Data Investment, Skills, and Firm Value. *Management Science*, *60*(6), 1452–1469.
Thomke, S. H. (2020). *Experimentation works: The surprising power of business experiments*. Harvard Business Press.
Van Maanen, J. E., & Schein, E. H. (1977). *Toward a theory of organizational socialization*.
Vinokurova, N., & Kapoor, R. (2020). Converting Inventions into Innovations in Large Firms: How Inventors at Xerox Navigated the Innovation Process to Commercialize Their Ideas. *Strategic Management Journal*, 1–29.
Wang, D. (2019). *How New Versions of Products Spread Differently Than Entirely New Products How New Versions of Products Spread Differently Than Entirely New Products*.
Wang, D., Pahnke, E. C., & McDonald, R. M. (2021). The Past Is Prologue? Venture-Capital Syndicates' Collaborative Experience and Start-Up Exits. *Academy of Management Journal*, *ja*.
Wooldridge, J. M. (2010). *Econometric analysis of cross section and panel data*. MIT press.
Wu, L., Hitt, L. M., & Lou, B. (2019). Data Analytics Skills, Innovation and Firm Productivity. *Management Science*, 1–41. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2744789
Wu, L., Lou, B., & Hitt, L. M. (2019). Data Analytics Supports Decentralized Innovation. *Management Science*.
Zbaracki, M. J. (1998). The Rhetoric and Reality of Total Quality Management. *Administrative Science Quarterly*, *43*(3), 602.
Zellweger, T. M., & Zenger, T. R. (2021). Entrepreneurs as scientists: A pragmatist approach to producing value out of uncertainty. *Academy of Management Review*, *October*.

Appendix

Appendix A. Additional details about sample construction

I began constructing my data sample by finding CPG companies in the Nielsen data large enough to have sufficient employee representation on the career networking website. As a first step, I merged firms identified in the Nielsen data with public firms in the Compustat database, which yielded 585 public companies. Though I do not limit my analysis to public firms, linking with Compustat was a quick way to identify relevant large CPG firms in the Nielsen data. It also served as an accuracy check: the number of linked firms I found is comparable to those of previous studies that matched Nielsen RMS data to Compustat (Argente et al., 2018). Using the company-name matching algorithm employed by Schoenle (2017), I first merged exact company-name matches. Then I eliminated common words and merged again. With the remaining names, I applied the "reclink" matching algorithm in Stata using the company name, address, city, and zip code, and kept matches with confidence scores greater than 85%. I manually examined all matches to check for accuracy.

I also manually identified 217 large private companies in the Nielsen data. I identified private companies by ranking the Nielsen companies by their number of distinct products in descending order. I manually checked the top 500 companies, and added them to the sample only if they had more than 50 online job reviews (i.e., a bare-minimum sample size for usability in the study). I stopped after manually checking the top 500 companies, having reached a saturation point below which companies no longer had over 50 online reviews.

Combining the public and private companies that I identified, I restricted the combined sample to the 451 firms that had ever had at least 50 distinct products (UPC codes) listed in the Nielsen data. I then identified consumer-product manufacturers by restricting the sample to companies in the 2-digit NAICS codes 31, 32, and 33. This step effectively excluded private-label retailers such as Wal-Mart and companies that sold non-CPG items such as software and gift cards. I then combined subsidiaries under public parent companies, aggregating employee review and Nielsen data to the same level as the public firm, resulting in 224 firms. Finally, to filter out firms whose online presence was too limited for meaningful text analysis, I restricted the sample to firm-years with at least 25 Glassdoor employee reviews, resulting in a sample of 154 firms.

I obtained résumé data from the networking website for employees who had worked at any of the 154 firms anytime between 2010 and 2016. Because my aim was to measure the culture of employees directly involved in product innovation, I extracted profiles only for employees whose job functions included "product management" or "research," using the networking website's internal job-function classification system. This procedure yielded 101,919 unique employee résumés in 182,403 positions at the firms in my sample. After restricting to firm-years with at least 100 job description texts, I was left with 61 firms and 376 firm-years (however, note that in the regressions, firm-years with no positive outcomes are automatically dropped from conditional logistic and Poisson models). This final sample represents a nearly comprehensive sample of large CPG firms that operate in the United States.

Appendix B. Method for measures derived from employee résumés

Constructing the measures derived from employee résumés (e.g., quantitative and qualitative analysis) proceeded in the following four steps:

Step 1. Identifying innovation-related sentences
To identify innovation-related sentences in employee résumés, I defined the dictionary of product-innovation related words by manually inspecting a random sample of 1000 job descriptions for words related to product innovation. As a guiding principle, I included words that were synonyms or explicitly linked to products (e.g. "product" and "launch"), innovation (e.g. "innovation" and "breakthrough"), or consumer market demand (e.g. "consumer" and "customer"). After creating an initial list of words, I searched the entire sample of job descriptions for sentences that contained at least one word from my initial list of words, and supplemented my initial list with product innovation-related words in those sentences. Finally, I searched the entire sample of sentences that contained words from my list, and pruned out words that counted irrelevant topics (e.g., specifying "product" or "products" must match a whole word, to avoid counting the word "production"). Innovation-related sentences are defined as sentences that match (using regular expressions) at least one word from the final "innovation" dictionary (see Table B1 for the words in the innovation dictionary).

Table B1. Words used to identify innovation-related sentences

Innovation Dictionary	"product\\b|products\\b|pipeline|launch|new brand|new \\S+ brand|new \\S+ \\S+ brand| new line|new \\S+ line|new \\S+ \\S+ line|new business|new \\S+ business|new \\S+ \\S+ business|new categor|new \\S+ categor|new \\S+ \\S+ categor|new opportun| transform\\w* \\S+ \\S+ categor|transform\\w* \\S+ categor|transform\\w* categor| growth|commercializ|innovat|invent\\b|invention|inventing|breakthrough|disrupt|radical| moonshot|moon shot|customer|consumer|user|market\\b|markets\\b"

Step 2. Training a word embedding model
I used the sample of innovation-relevant sentences from employee job descriptions to train a 300-dimensional word embedding model, using the same method described in Li et al. 2021. The Li et al. 2021 method involves providing seed words to the word embedding model, in order to ultimately build a relevant dictionary of words. For example, seed words for quantitative analysis included "data analytics" and "quantitative". The word embedding model then creates a dictionary of related words. I identified seed words using the same process as described above for the innovation dictionary—manually inspecting a random sample of 1000 job descriptions, iteratively adding and pruning words to create a parsimonious list. The specific seed words used for each word vector are displayed in Table B1.

Table B2. Seed words used in Word Embedding Model

Quantitative Seed Words	\\bdata|\\bquant\\b|\\bquantitative|\\banalytics|\\bpredictive analytic|\\bstatistic| \\bbusiness intelligence|\\bbi\\b|\\bdashboard|\\bomniture|\\balteryx|\\bminitab|\\bcognos| \\bjmp\\b|\\bgis\\b|\\bqlik|\\blooker\\b|\\bpowerbi|\\bdomo\\b|\\btableau|\\besri\\b|

	\\bsisense|\\byellowfin|\\bsas visual analytic|\\bspotfire|\\bmicrostrategy|\\boracle bi| \\boracle analytic|\\bregress|\\bsurvival analysis|\\bcausal inference|\\bcausal estimate| \\bestimate causal|\\bcausal effect|\\bcausal impact|\\bmultivariate| \\bmulti-variate|\\beconometric|\\btime series|\\bpanel data|\\bspss|\\bsas\\b|\\bstata| \\bmachine learn|\\bmachine-learn|\\bml\\b|\\bai\\b|\\bartificial intelligence| \\brandom forest|\\bdecision tree|\\bsupport vector machine|\\bsvm\\b|\\bnnet| \\bneural net|\\bdeep learn|\\bclustering|\\bcluster analy| \\bk mean|\\bk-mean|\\bsentiment analy| \\bnlp\\b|\\bnatural language process|\\btext analy| \\b(?:(mining|mine|mined)\\W+(?:\\w+\\W+){0,3}?(data)| (data)\\W+(?:\\w+\\W+){0,3}?(mining|mine|mined))\\b|\\btensorflow|\\bhadoop| \\bpytorch|\\brapidminer|\\bbigquery|\\bsnowflake|\\bpython|\\bmatlab|\\bpandas| \\bredshift|\\bnumpy|\\bscikit|\\bjupyter|\\bnielsen|\\bgfk\\b|\\bkantar\\b|\\biri\\b| \\bconsumer panel|\\bscanner data|\\bhousehold panel| \\blifetime customer value|\\bfactor analys|\\bdiscrete choice analy| \\bdca\\b|\\bconjoint|\\bbases test|\\ba/b\\b|\\bab test| \\boptimizely|\\bsurvey|\\bquestionnaire|\\bpolled|\\bpolling|\\bnps\\b| \\bnet promoter score\\b|\\bqualtric|\\bsurveymonkey|\\bexcel\\b|\\bspreadsheet| \\bvba|\\bsql|\\bpowerpivot
Qualitative Seed Words	\\bqualitative|\\bqual\\b|\\bfocus group|\\bethnograph|\\banthropolog|\\binterview| \\bf2f\\b|\\bface-to-face|\\bshop-a-long|\\bstory|\\bstories|\\bnarrat| \\bconsumer journey|\\buser journey|\\bcustomer journey|\\bjourney map| \\bexperience journey|\\bpersona\\b|\\bpersonas\\b|\\bcase stud| \\bepics\\b|\\bepic\\b| \\b(?:(customer\\w*|consumer\\w*|user\\w*|shopper\\w*|\\bhome\\w*) \\W+(?:\\w+\\W+){0,5}? (interact\\w*|conversation\\w*|observ\\w*|talk\\w*|visit\\w*)| (interact\\w*|conversation\\w*|observ\\w*|talk\\w*|interfac\\w*|visit\\w*) \\W+(?:\\w+\\W+){0,5}?

	(customer\\w*|consumer\\w*|user\\w*|shopper\\w*|\\bhome\\w*))\\b
Early Stage Seed Words	\\binductive\\w*|\\bfound\\b|\\bfind\\w*|\\breveal\\w*|\\bdiscover\\w*| \\blearn\\w*|\\bunderstand\\w*|\\bunderstood\\b|\\bempath\\w*|\\bdetect\\w*| \\bobserv\\w*|\\bprobe\\b|\\bprobing\\b|\\bsearch\\w*|\\bexplor\\w*| \\binterpret\\w*|\\bidentif\\w*|\\binvestigat\\w*|\\bexamin\\w*|\\bdive\\b| \\bdove\\b|\\bdig\\b|\\bdug\\b|\\bdelve\\w*
Late Stage Seed Words	\\bdeduc\\w*|\\bhypothesi\\w*|\\bexperiment\\w*|\\bevaluat\\w*| \\bvalidat\\w*|\\bverif\\w*|\\bprove\\w*|\\bproof\\b|\\bconfirm\\w*| \\bevidence\\w*|\\bdemonstrat\\w*|\\bjustif\\w*|\\btest\\w*|\\bpilot\\b| \\btrial\\b|\\btrials\\b

Step 3. Building a final word dictionary

The word embedding model identified about 500 relevant words for each measure. I manually pruned unrelated words to create a final word dictionary for each of my measures (K. Li et al., 2021). For the qualitative measure, I added an additional step. Unlike quantitative analysis (e.g. “data analytics”) many instances of qualitative analysis were not easily identifiable using standalone terms. Rather, employees would provide a description of interviewing a customer or observing a field site (e.g., “conducted interviews to determine customer preferences”). Thus, I added to the qualitative dictionary any instance that matched the following regular expression:

```
\\b(?:(customer\\w*|consumer\\w*|use\\w*|shopper\\w*|\\bhome\\w*|\\bfield|\\bsite)\\W+(
?:\\w+\\W+){0,5}?(interview\\w*|interact\\w*|conversation\\w*|observ\\w*|talk\\w*|imm
ers\\w*|visit\\w*)|(interview\\w*|interact\\w*|conversation\\w*|observ\\w*|talk\\w*|interf
ac\\w*|immers\\w*|visit\\w*)\\W+(?:\\w+\\W+){0,5}?(customer\\w*|consumer\\w*|use\\
w*|shopper\\w*|\\bhome\\w*|\\bfield|\\bsite\\w*))\\b
```

This regular expression finds customer-related nouns (e.g., customer, consumer) within five words of relevant verbs (e.g., interview, observe) or vice versa.

Table B3 lists all the words used for each measure, listed from most frequent to least frequently occurring (see also Table 1 in the paper):

Table B3. Final Word Dictionaries

Final Quantitative Dictionary	data, analytics, quantitative, survey, bi, statistic, dashboard, data analysis, nielsen, business intelligence, market data, sql, customer data, data driven, big data, data analytics, consumer data, tableau, sales data, iri, nps, data science, regress, concept test, gfk, consumer panel, net promoter score, predictive analytic, syndicated data, kantar, qlik, spreadsheet, vba, questionnaire, clustering, cognos, microstrategy, conjoint, panel data, omniture, spotfire, nlp, gis, minitab, multivariate, quant, microsoft excel, household panel, text analy, jmp, time series, ab test, powerbi, sentiment analy, econometric, natural language process, cluster analy, alteryx, factor analys, bases test, lifetime customer value, oracle bi, powerpivot, qualtric, scanner data, sisense, survival analysis, domo, esri, looker, polled

Final Qualitative Dictionary	*[interactions/visits/interview, etc.] within 5 words of [customer/consumer/field, etc.]*, qualitative, user stories, user acceptance test, focus group, use case, voice of customer, voc, uat, voice of the customer, customer journey, acceptance criteria, ethnograph, usability test, personas, case stud, user story, face to face, persona, consumer journey, user journey, epics, storyboard, qual, one on one, epic, journey map, field research, anthropolog, experience journey, field stud
Final Early Stage Dictionary	discovery, front end innovation, unmet, exploration, detection, explore, discover, early stage, exploring, uncover, exploratory, explored, detect, uncovered, uncovering, explorer, discovered, discovering, detecting, detector, early stages, explorations, fuzzy front end, detected, reveal, discovers, revealed, detectors, discoveries, detects, discoverability, revealing, detective, discoverable, discoverer, discovery, explorative, front end innovations, reveals, detectable, discoverant, explorers, explores, front end innovator, inductive, uncovers
Final Late Stage Dictionary	evaluation, prototyp, validation, pilot, trials, evaluate, trial, evaluated, selection, consumer test, verification, evaluating, validate, selected, evaluations, validated, select, screening, validating, validations, verify, screen, market test, late stage, verified, selecting, verifying, confirm, development stage, evaluates, screens, customer test, selective, confirmed, confirming, hut, confirmation, go no go, screened, selections, validates, confirmatory, screenings, commercialization phase, verifications, evaluator, screener, verifies, evaluative, selects, screenshots, trialing, verifire, confirms, evaluators, screeners, selector, confirmations, screenplays, screensaver, screenshot, selectability, selectively, trialed, validator, verifeye, verifing, verifiying

Step 4. Aggregating to create the final measures

I created the final measures by counting words in the corresponding word dictionary (see Table B3). I defined each measure as the mean number of words in each innovation-related sentence.

Appendix C. Additional Validation of Quantitative and Qualitative Analysis Measures

Validation of Quantitative Analysis Measure with External Data Source: Earnings Calls
To validate the quantitative analysis measure, I obtained earnings call transcript data from Wharton Research Data Services database for all the public firms in my sample. (I did not use these data to validate the qualitative measure because there were very few mentions of qualitative words in earnings calls). Using only "Executives" speaker types (i.e., not including text from analysts external to the company), I created the measure *Earnings Calls Quantitative* counting the total number of words (identified using the quantitative word dictionary displayed in Appendix Table B3), and *Earnings Calls Quantitative %* by dividing that measure by the total number of words. As a control, I also created the measure *Earnings Calls Innovation %* by dividing the number of innovation-related words (using the innovation word dictionary in Appendix Table B1) by the total number of words.

In Table C1, I show that *Earnings Calls Quantitative* and *Earnings Calls Quantitative %* has significant positive between and within-firm correlation to my primary *Quantitative Analysis* measure. Column 1 regresses *Quantitative Analysis* on *Earnings Calls Quantitative*, indicating a positive significant relationship ($p<0.05$). Columns 2-4 demonstrate that the relationship is robust to company and year fixed effects, and additional controls. Column 5 shows the relationship holds using the measure *Earnings Calls Quantitative %.*

Table C1. Validation with Independent Data Source: Earnings Call Transcripts

Dependent Variable: Quantitative Analysis$_{ft}$	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**
Earnings Calls Quantitative$_{ft}$	0.006^{***}	0.002^{*}	0.127^{**}	0.135^{*}	
	(0.002)	(0.001)	(0.046)	(0.058)	
Earnings Calls Quantitative %$_{ft}$					0.023^{*}
					(0.011)
Earnings Calls Word Count$_{ft}$			-0.009	-0.009	0.037
			(0.045)	(0.042)	(0.043)
Company Controls	NO	NO	NO	YES	YES
Company FE	NO	YES	YES	YES	YES
Year FE	NO	YES	YES	YES	YES
Adj. R^2	0.091	0.818	0.817	0.816	0.813
Within R^2		0.022	0.023	0.033	0.016
N	316	316	316	316	316

Notes: coefficients are estimated using OLS regression on firm-year panel data (subscript *f* indicates firm-level and *t* indicates year-level variation). Standard errors (in parentheses) are clustered at the firm level. $^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$; $^{+}p < 0.1$

Individual-level Analysis
As another validation of the quantitative and qualitative analysis measures, I demonstrate that individuals conformed to the organizational culture they joined: when an individual moved from an organization that uses less quantitative (or qualitative) analysis to one that uses more, his or her own individual use of that type of analysis significantly increases.

Using individual-level data from the online career networking website, I constructed two datasets: 1) a Job Description Dataset with one observation for each distinct job description; and 2) a Mobility Dataset with one observation for each time an individual moved from one of the companies in my sample to another company in my sample (i.e., companies for which I had measures of firm-level quantitative and qualitative analysis). Both datasets were restricted to the 61 companies that were used in the final analysis in the paper (which is why there are only 50,876 job descriptions in the Job Description dataset used for Table C2). In each dataset, I used the same methodology described in the paper to construct variables for quantitative and qualitative analysis at the individual and company level: *Individual Quantitative, Individual Qualitative*, *Company Quantitative*, and *Company Qualitative.* For each individual, I adjusted the company-level measures to be the mean of all individuals of the company *except* the individual's own value. This adjustment avoided a mechanical relationship between individual and company-level measures that would necessarily arise if the individual were counted as part of the company. I also Winsorize at the 5% level due to extreme outliers.

Table C2, columns 1-4 use the larger Job Description-level Dataset. Column 1 shows that *Company Quantitative* has the largest and most significant effect on *Individual Quantitative*, and Column 4 shows that *Company Qualitative* has the largest and most significant effect on *Individual Qualitative*. Columns 5 and 6 repeat columns 3 and 4, adding a control for formal statistical education. The effects hold, and, as expected, formal statistics education has a much larger effect predicting *Individual Quantitative* than *Individual Qualitative*.

Table C2, columns 5 and 6 use the Mobility Dataset. In column 5 I regress *ΔIndividual Quantitative* (the difference between the individual's new and previous quantitative analysis measure) on *ΔCompany Quantitative* (the difference between the new and previous company's quantitative analysis measure). Column 6 does the same for qualitative analysis. Both regressions show that when an individual moved from an organization that uses less quantitative (or qualitative) analysis to one that uses more, his or her own individual use of that type of analysis significantly increases.

Table C2. Individual-level Analysis

	(1) DV: Individual Quant	**(2) DV: Individual Qual**	**(3) DV: Individual Quant**	**(4) DV: Individual Qual**	**(5) DV: ΔIndividual Quant**	**(6) DV: ΔIndividual Qual**
Company Quant	0.506***	0.010***	0.491***	0.010***		
	(0.021)	(0.003)	(0.021)	(0.003)		
Company Qual	0.254***	0.168***	0.243***	0.168***		
	(0.068)	(0.009)	(0.068)	(0.009)		
Formal Stats Edu			0.114***	0.001+		
			(0.006)	(0.001)		
Δ Company Quant					0.231**	0.020+
					(0.085)	(0.011)
Δ Company Qual					0.048	0.114**
					(0.301)	(0.039)
Adj. R^2	0.014	0.009	0.022	0.009	0.002	0.005
N	50876	50876	50876	50876	2407	2407

Appendix D. Robustness Checks

Table D1. Robustness Checks

	OLS DV: New Product Sales				Logistic DV: Breakthrough			
	(1) Discrete Qual	(2) Winsor-ized	(3) Placebos	(4) Public Controls	(5) Winsor-ized	(6) Placebos	(7) Public Controls	(8) NPCBPS Weights
Quant Analysis	-0.207	0.086	0.075	0.008				
	(0.237)	(0.167)	(0.166)	(0.204)				
Qual Analysis		0.108	0.110	0.025				
		(0.191)	(0.189)	(0.269)				
Quant Analysis × Qual Analysis		0.388**	0.424**	0.346*				
		(0.129)	(0.125)	(0.157)				
Qual Analysis (2nd Tercile)	0.180							
	(0.212)							
Qual Analysis (3rd Tercile)	0.278							
	(0.330)							
Qual Analysis (2nd Tercile) × Quant Analysis	0.163							
	(0.240)							
Qual Analysis (3rd Tercile) × Quant Analysis	0.770**							
	(0.285)							
Percent Quantitative					0.010	-0.011	-0.238	-0.142
					(0.117)	(0.124)	(0.183)	(0.134)
Percent Quantitative Squared					-0.248*	-0.296***	-0.326**	-0.278*
					(0.101)	(0.083)	(0.105)	(0.119)
Number of Profiles	0.170	0.543	0.996	2.664^{+}	-0.875*	-0.346	0.581	-1.774***
	(0.734)	(0.683)	(0.791)	(1.436)	(0.412)	(0.549)	(1.141)	(0.491)
Product Introductions (lag)	-0.460^{+}	-0.447^{+}	-0.396	-0.020	-0.341*	-0.322*	-0.119	-0.422*
	(0.242)	(0.248)	(0.247)	(0.232)	(0.163)	(0.150)	(0.139)	(0.188)
Total Sales	-0.062	0.269	-0.054	-0.702**	0.315	0.381	0.936	0.873
	(0.154)	(0.605)	(0.165)	(0.235)	(0.501)	(0.353)	(0.667)	(0.711)
Total Products	0.365	0.359	0.962	1.828	0.063	0.745	2.811**	0.778
	(0.797)	(0.884)	(0.860)	(1.184)	(0.648)	(0.678)	(1.025)	(0.897)
Early Stage			0.109			0.011		
			(0.240)			(0.142)		
Late Stage			0.015			0.159		
			(0.161)			(0.132)		
Quant Skills			-0.399			-0.476*		
			(0.274)			(0.229)		
Qual Skills			-0.041			0.165		
			(0.247)			(0.324)		
Years of Quantitative Experience			-0.248			-0.273		
			(0.368)			(0.345)		

	OLS DV: New Product Sales				**Logistic DV: Breakthrough**			
	(1) Discrete Qual	**(2) Winsor-ized**	**(3) Placebos**	**(4) Public Controls**	**(5) Winsor-ized**	**(6) Placebos**	**(7) Public Controls**	**(8) NPCBPS Weights**
Years of Formal Quantitative Education			-0.150			0.024		
			(0.120)			(0.113)		
Early Stage × Late Stage			-0.109			-0.086		
			(0.139)			(0.076)		
Quant Skills × Qual Skills			0.055			0.039		
			(0.180)			(0.173)		
Price				0.002			-0.002	
				(0.003)			(0.002)	
Co. Market Value				0.000**			0.000*	
				(0.000)			(0.000)	
Co. ROA				3.438			-6.589	
				(11.254)			(7.356)	
Co. Net Income				0.000			0.000***	
				(0.000)			(0.000)	
Co. % Research Job Titles (Glassdoor)				-0.539			1.429	
				(3.150)			(1.656)	
Co. % Data Job Titles (Glassdoor)				-8.116			-8.557+	
				(6.489)			(5.110)	
Co. % Data Job Titles (Online Profiles)				22.299			18.341	
				(15.420)			(11.764)	
Glassdoor Overall Rating				-0.260			0.197	
				(0.523)			(0.459)	
Glassdoor Culture Rating				-0.034			0.018	
				(0.278)			(0.265)	
Co. R&D Expenditure				-0.001			-0.000	
				(0.000)			(0.000)	
Co. Employees				0.009			0.009	
				(0.017)			(0.006)	
Co. CapEx				0.000			-0.000	
				(0.000)			(0.000)	
Company FE	YES	YES	YES	YES	YES	YES	YES	YES
Year FE	YES	YES	YES	YES	YES	YES	YES	YES
Product Group FE	YES	YES	YES	YES	YES	YES	YES	YES
Adj. R^2	0.117	0.118	0.117	0.099				
Within R^2	0.006	0.006	0.007	0.012				
Pseudo R^2					0.115	0.118	0.140	0.138
N	3502	3502	3502	2024	3326	3326	1869	3052

***$p < 0.001$; **$p < 0.01$; *$p < 0.05$; +$p < 0.1$

Figure D1: Covariate Balance for NPCBPS weighting (used in Table D1 column 8)

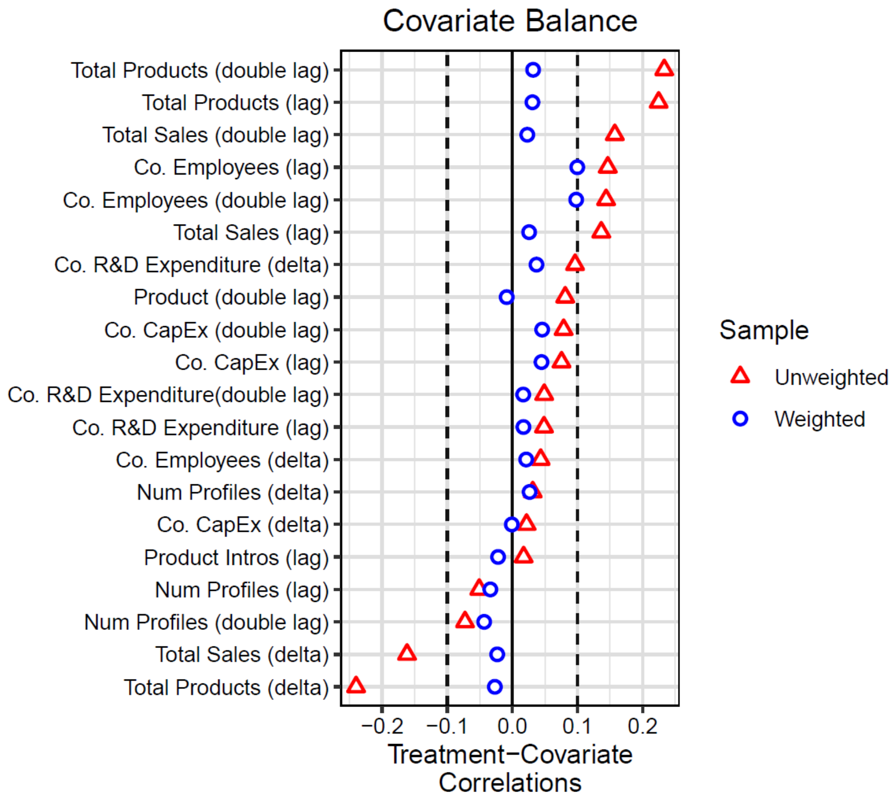


Table D2. Novel Breakthroughs vs. Other Breakthroughs

	Logistic; Product-level data		Poisson; Firm panel data	
	(1) DV: Novel Breakthrough	**(2) DV: Other Breakthrough**	**(3) DV: Novel Breakthroughs**	**(4) DV: Other Breakthroughs**
Quantitative Analysis$_{ft-1}$	0.113	-0.182	0.106	-0.148
	(0.167)	(0.240)	(0.215)	(0.221)
Qualitative Analysis$_{ft-1}$	-0.217	-0.299^{+}	-0.164	-0.192
	(0.202)	(0.169)	(0.245)	(0.175)
Quant Analysis$_{ft-1}$ × Qual Analysis$_{ft-1}$	0.405*	0.178	0.459*	0.207
	(0.177)	(0.148)	(0.205)	(0.152)
Company Controls	YES	YES	YES	YES
Company FE	YES	YES	YES	YES
Product Group FE	YES	YES	NA	NA
Year FE	YES	YES	YES	YES
Pseudo R^2	0.145	0.136	0.263	0.398
N	1822	2677	199	204

Notes: Coefficients in columns 1-2 were estimated using logistic regression on product-level data, columns 3-4 were estimated using Poisson regression on firm-year panel data. Control variables are included, as in Table 4. Robust standard errors are clustered at the firm level. $^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$; $^{+}p < 0.1$

Chapter 2. Market Size Inversion: How Diffusion Dynamics Invert Market Size Expectations for Novel Products

1. Introduction

There are many popular anecdotes of novel innovations that appeared to have low demand prior to launch, but ultimately achieved breakthrough success. Apple's wireless AirPod headphones provide one example.[19] Following their 2016 release, critics mocked the AirPods' design, noting that "they look... just like the old EarPods, with the wires cut off", hinting that the price didn't justify the new features, wondering "will they stay in your ears?" and asking "would people actually wear these?" (Kelly 2016, Patel 2016). But by 2018, the product had become a status symbol, topping the wireless hearables category with 35 million units sold that year, and even facing supply issues due to unexpectedly high customer demand (Fingas 2018, Miller 2019).

Yet inversely, many seemingly safe non-novel products fail despite high forecasted demand. According to some estimates, up to 75% of new consumer products fail, despite millions of dollars invested into market research (Schneider and Hall 2011). Given that no firm intends to launch an unsuccessful product, these products likely underperformed initial market size expectations.

These observations highlight what I call the "market size inversion" puzzle: Why do many breakthrough products achieve success despite low anticipated demand, while many other products fail despite high anticipated demand? A growing body of strategy research provides a possible explanation: novel products appear less commercially attractive due to high demand uncertainty, which can arise from customers' lack of understanding how new features benefit them (Anthony

[19] As another example, Airbnb CEO Brian Chesky, reflecting on early pushback to the business, remarked that "all these really good ideas or big ideas often sound like stupid ideas. Somebody once told me in the early days, don't worry about anyone stealing your idea. If it's any good, everyone will dismiss it." (Chesky, 2015; Zellweger and Zenger, 2021). For more examples, see Felin and Zenger (2017, 2018)

et al. 2016, von Hippel 1986); from unidentified future uses (Bresnahan 2010, Rosenberg 2009); from the evolving trajectories of price points and functional needs (Adner and Levinthal 2001, Adner and Snow 2010, Christensen and Bower 1996); and from a lack of institutional support (Gao and McDonald 2022, Ozcan and Gurses 2018). Despite appearing unattractive, such highly uncertain products actually offer the highest potential for innovative rents, because the rents for obvious (non-novel) opportunities are competed away (Denrell et al. 2003). The best innovative opportunities are captured by those who have a superior cognitive ability to perceive how valuable these opportunities will be, despite appearing unattractive in the present (Felin and Zenger 2017, Gavetti 2012). Therefore, absent other competitive barriers to entry, the most successful innovations are almost by definition initially surprising or contrarian (Felin et al. 2019, Zellweger and Zenger 2021). In short, the prevailing explanation is that demand uncertainty (about features, uses, regulation, price points, etc.) is a competitive barrier to entry, which provides valuable opportunities for uncontested innovative rents for innovators who have superior cognitive ability to perceive these opportunities.

Although deeply insightful, unresolved issues persist with this existing competition-based explanation of market size inversion. One issue is that the theory only provides a *post hoc* explanation of surprising breakthrough commercial successes. Looking backward, the theory identifies that being surprising or contrarian is practically a pre-condition for breakthroughs, but is not able to discern which products will be breakthroughs in the future. Relatedly, according to the theory, it is unclear when conducting additional market research for a new product would be helpful (by increasing the innovator’s ability to perceive new opportunities) or harmful (by anchoring on what people already know they like). In this paper, I suggest that resolving these issues requires not only understanding supply-side factors (e.g., competition or innovator cognition), but also an integration of the demand-side dynamics of how products diffuse via social processes (Abrahamson and Rosenkopf 1997, Rogers 1962). By explicitly accounting for how diffusion dynamics change depending on product attributes (novelty), this paper makes progress toward *a priori* modeling which specific products may defy market size expectations, and for understanding when to rely less on market research.

In my model, customers' evaluations of relatively novel products depend more on neighbors' endorsements than on their own individual evaluations. But neighbors' endorsements do not exist until the product diffuses. Therefore, for relatively novel products, a larger portion of potential demand does not exist until after the product diffuses, thereby downward biasing estimates of market demand prior to launch. Exacerbating the problem, for less novel products, customers are more certain of their own individual evaluations, but the product may socially diffuse at a slower rate, thus leading to upward biased estimates of market demand. Taken together, these forces can lead to a systematic inversion of market size expectations. I illustrate these dynamics using an agent-based simulation model (code to recreate the model will be available in the online supplementary materials).

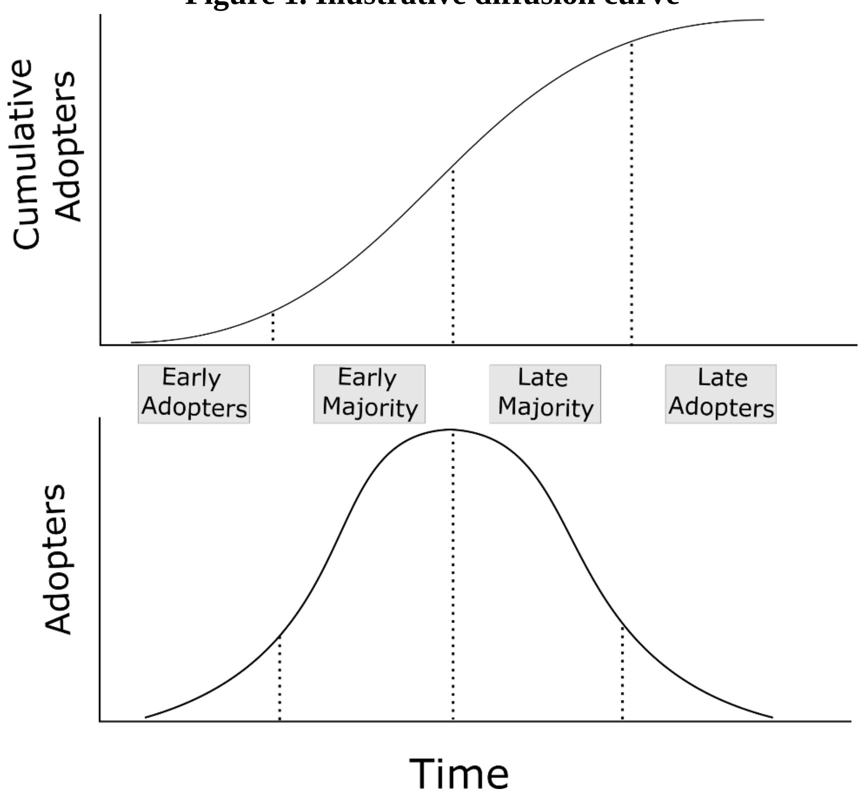
I empirically validate several assumptions and predictions of the model using a dataset of product-level diffusion for about 1,600 products launched by firms in the Consumer-Packaged Goods (CPG) industry from 2010-2016. First, I validate my model's assumption that customers rely more on neighbors' evaluations for novel products, by demonstrating that diffusion for novel products is more likely to fit a socially-driven adoption curve (Bass 1969). And second, I validate key implications of the model by combining the product data with firm résumé data. Using this combined dataset, I show that firms that rely more on quantitative market sizing are indeed more likely to fall into what I call the "market size trap": they are less likely to produce highly novel products, and less likely to produce breakthrough products, because these appear less attractive when relying on quantitative market sizing.

This paper makes two contributions to the strategy literature. The primary contribution is a demand-side model which explains market size inversion (i.e., novel products overperform expectations while non-novel products underperform expectations), purely as a function of diffusion dynamics. Although my diffusion-based model can stand alone as an alternative explanation to the existing competition-based model of demand uncertainty as a barrier to entry (Denrell et al. 2003, Felin and Zenger 2017, Gavetti 2012), the two models are best viewed as complements. My diffusion-based model aids the prior competition-based model in discerning which products will be successful in the future (not just *post hoc*), and when market research about customer preferences is useful vs. misleading. The paper also makes a secondary contribution to the research stream on resolving uncertainty in nascent technologies. Whereas prior work in this stream has inherently taken a technological (Anthony et al. 2016, von Hippel 1986, Rosenberg 2009) or market-oriented (Adner and Levinthal 2001, Gao and McDonald 2022) view of demand uncertainty, this paper illuminates additional uncertainty that arises from customers when viewing adoption as social diffusion process. Thus, some uncertainty for nascent technologies may be resolvable by better understanding and accounting for the demand-side dynamics modeled in this paper.

2. Innovation Diffusion

My paper aims to explain market size inversion (novel products overperform, non-novel products underperform) as a function of social diffusion dynamics. According to theories of innovation diffusion, the majority of potential customers in a market decide to adopt an innovation only after observing others' adoption (Geroski 2000, Moore 1991, Valente and Rogers 1995). This social diffusion process is especially likely to apply to novel innovations. Because novel innovations are by definition different from what existed before, the majority of customers in the market are uncertain whether new innovation is valuable to them. But a small group of early adopters—those with specific needs met by the innovation, or a greater tolerance for ambiguity—adopt the new innovation despite the uncertainty and risk associated with it. As the early adopters demonstrate the value of the new innovation, the early majority becomes willing to try, followed by the late majority, and so on until the innovation diffuses through the population (Rogers 1962).[20] Figure 1 shows an illustrative depiction of how this staged social diffusion process gives rise to an S-shaped cumulative adoption curve over time.

Figure 1. Illustrative diffusion curve



[20] Technically, Rogers' (1962) seminal work conceptualizes five categories of individuals, partitioned by standard deviations of a normal distribution from the individuals' mean adoption time: innovators who adopt first (2.5% of adopters), early adopters (13.5%), early majority (34%), late majority (34%), and laggards who adopt last (16%).

But not all innovation diffusion processes are socially driven to the same extent. Ambiguity about the value of the innovation determines the degree to which it socially diffuses. Due to never-before-seen features or value propositions, novel innovations are more ambiguous to evaluate, and therefore tend to follow the social diffusion process described above (Bass 1969). Under the ambiguity of novelty, most customers wait until observing neighbors' evaluations of the innovation before adopting themselves. But non-novel innovations are less ambiguous—people know what to expect, and can rely less on neighbors' evaluations before deciding to adopt. Because the adoption decisions are less socially driven, the adoption curves of non-novel innovations may not exhibit the same S-shaped cumulative adoption. For example, replacement products (e.g., the iPhone 13 vs. the iPhone 10) are characterized by less ambiguity than entirely new products. Rather than following a socially-driven S-shaped cumulative diffusion curve, they diffuse along a curve that resembles a power law distribution (Jin et al. 2019, Wang 2019).

Because novelty affects the degree to which an innovation socially diffuses, prior work has conceptualized potential customer's adoption decision as a function of two distinct components: 1) individual assessment based on personal preferences; and 2) assessment based on neighbor endorsement (Abrahamson and Rosenkopf 1997, Banerjee et al. 2013). When utility for an innovation is relatively ambiguous (as is the case with relatively novel products), potential customers rely more on neighbors' evaluations relative to their own individual preferences, leading to a more socially-driven diffusion process.

A central argument of this paper is that because customers' adoption decisions depend relatively more on neighbors' endorsements for novel products, a larger portion of estimated demand does not exist until after the diffusion process. Therefore, relatively novel products will have downward biased market size estimates prior to diffusion. The next section will formalize this argument, using an agent-based model to illustrate.

3. Model Structure

In this section, I develop an agent-based model simulate the extent to which an innovation diffuses among potential adopters. I will use the model to explore how varying two parameters (initial customer assessments and novelty) affects the expected market size and the actual extent of diffusion. Each simulation evaluates an innovation's extent of diffusion among 100 potential adopters. [21] The simulation progresses by discrete time periods. In the most basic formulation of the simulation, all potential adopters are aware of the innovation, and so decide whether to adopt the innovation each time period (this assumption of perfect information will be relaxed later in the paper). They adopt in a given time period if their overall assessment of the innovation is positive. The simulation stops there are no new adopters.

Following prior work (Abrahamson and Rosenkopf 1997), I model each potential adopter's overall assessment of the innovation as a function of two parts: individual assessment and neighbor adoption. I express this with the equation,

$$Y_{ijt} = \left(1 - N_j\right) \times I_i + N_j \times P_{t-1}$$

Where $\boldsymbol{Y_{ijt}}$ is potential adopter i's **overall assessment** of the innovation j in period t; $\boldsymbol{I_i}$ is the adopter's **individual assessment** of the innovation; $\boldsymbol{P_{t-1}}$ is the **proportion of neighbors' adoption** of the innovation in period $t-1$; and $\boldsymbol{N_j}$ is the **novelty** of the innovation.[22] In a given time period, the individual adopts the innovation if $Y_{ijt} > 0$. I define the anticipated market size

[21] Whereas most early innovation studies focused on the rate of diffusion among the set of eventual adopters (Rogers 1962), more recent work focuses on the extent to which an innovation penetrates a given population (Abrahamson and Rosenkopf 1997). In this paper, I focus on the latter.

[22] Prior work has used the term *Ambiguity* rather than *Novelty*. Although *Ambiguity* is a more general term that would work in my model as well, I use the term *Novelty* for clarity to match the theorizing and the empirical results throughout the paper.

as the proportion of customers that have an initial positive individual assessment prior to the diffusion simulation (I_i>0), and the final extent of diffusion as the proportion with a positive overall assessment at the end of the diffusion simulation ($Y_{ijt} > 0$).

The basic intuition of the model is that for a more novel innovation (higher N_j), a potential adopter's overall assessment will weight neighbor adoption (P_{t-1}) relatively more than their individual assessment (I_i). Novelty, N_j, is one of the primary parameters of interest that will vary across simulations.

Another important model parameter is the mean of the distribution of potential adopters' individual assessments. Each potential adopter's individual assessment I_i is drawn from a normal distribution,

$$I_i \sim N(\mu, 0.1)$$

Where mean μ is a parameter of interest, and the standard deviation held constant at 0.1. The individual assessment for each potential adopter is drawn once at the beginning of the simulation and held constant throughout. In the analysis, the simulations will explore how the extent of innovation diffusion changes for different mean individual assessments (μ).

Proportion of neighbors adopted, P_{t-1}, is dynamically calculated throughout the simulation. In the most basic model, P_{t-1} is calculated as the proportion of the potential adopter's neighbors that had adopted the innovation in the prior period.

$$P_{t-1} = \frac{Adopter\ Neighbors_{t-1}}{Total\ Neighbors}$$

Each potential adopter's neighbors are constant throughout the simulation. Neighbors are assigned at the beginning of the simulation based on network ties to other potential adopters, as determined by a random Barabasi-Albert network. This is a common network structure for simulating the preferential attachment structure that is common in real social networks (Albert and Barabási 2002, Delre et al. 2007).[23]

In the next section, I will use this model to show that potential customers' adoption decisions depend more on neighbors' endorsements for relatively novel products, so a larger portion of demand does not exist until after the diffusion process. I will also present modified models that relax several assumptions in this baseline formulation of the model.

4. Model Analysis

The purpose of the simulation model is to explore how varying two parameters (novelty N_j and the mean μ of individual assessments I_i) impacts the anticipated market size and the actual extent of adoption for an innovation. I will build model intuition, and consider generalizability to different contexts, by presenting three sets of analyses based on increasingly complex versions of the simulation: *Model 1 - Simplified Baseline Simulation* as the analysis of the baseline formulation of the simulation model as described above; *Model 2 - Main Simulation* as analysis of the main simulation model of interest, which relaxes the assumption of perfect information among potential adopters; and *Model 3 - Negative Endorsements Modification* which adds addition of the possibility of negative neighbor endorsements that limit the extent of diffusion.

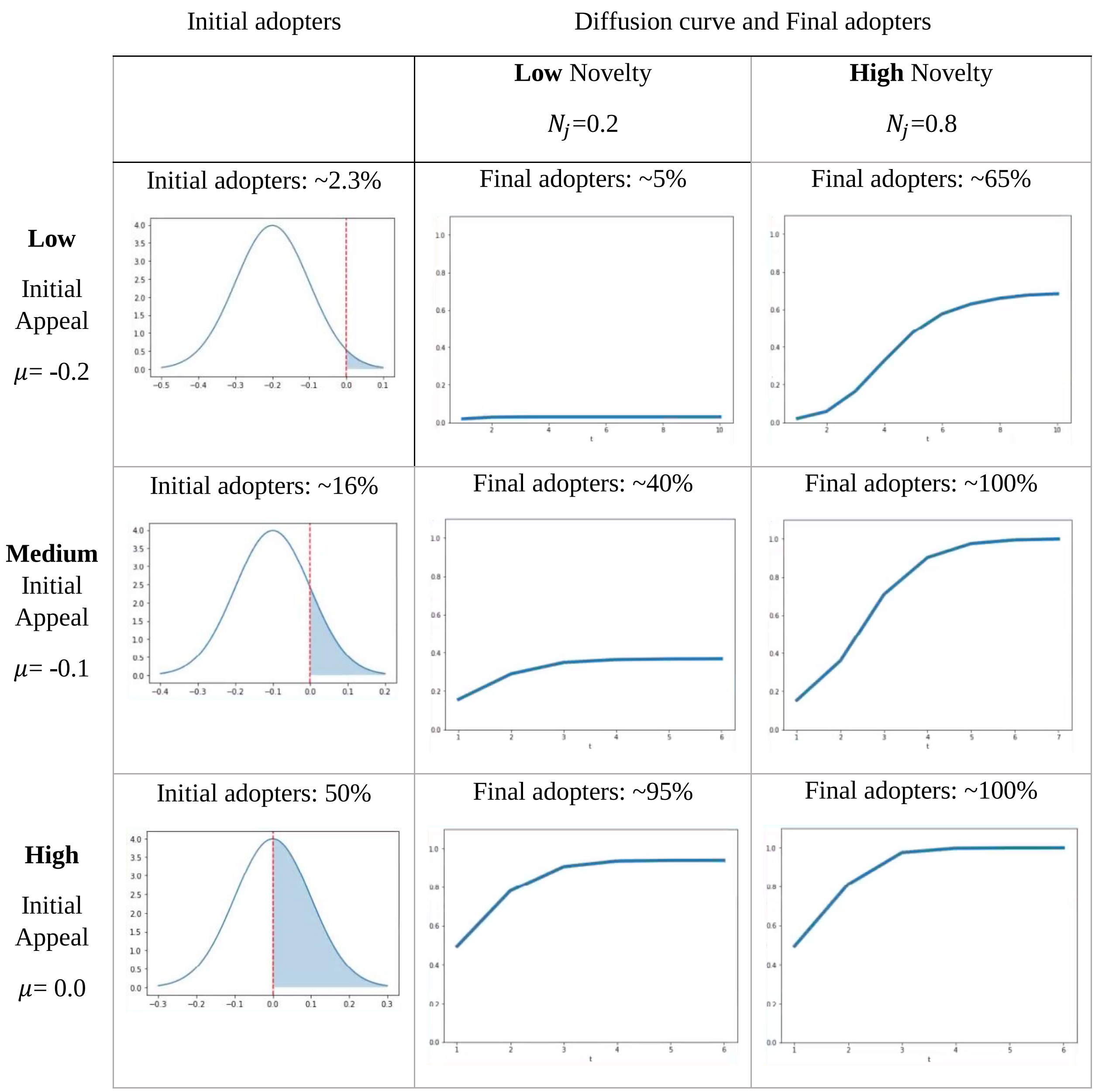
Model 1: Simplified Baseline Simulation

First, I will present several examples using a simplified baseline simulation to build intuition in the model. For illustrative purposes, Figure 2 displays the distribution of initial individual assessments and diffusion curves for several combinations of parameter values N_j and μ. The second row of Figure 2 presents an innovation with a medium level of potential adopters'

[23] The key implications of the model appear robust to the choice of the specific social network structure used. Network structure is not the focus of this paper, but could be a topic of future research.

initial assessment ($\mu = -0.1$). With this distribution of individual assessments, only about 16% of potential adopters have a high enough individual assessment of the innovation (>0) to adopt initially. But, after the diffusion process, the number who ultimately adopt increases dramatically: 40% ultimately adopt when the innovation has low novelty ($N_j = 0.2$), and nearly 100% adopt the innovation with high novelty ($N_j = 0.8$).

Figure 2. Baseline simulation examples



In this simple baseline version of the simulation, there is a stark difference in final adoption between low and high novelty innovations, especially for products with lower initial appeal. In the first row of Figure 2, the high novelty innovation diffuses to about 70%, while the low novelty innovation barely makes 5%. By contrast, there is little difference between high and low novelty in the third row of Figure 2.

Figure 3 makes this pattern even more clear. Rather than plotting the diffusion curves as in Figure 2, Figure 3 plots the extent of final adoption for various combinations of the parameters. The x-axis plots the initial estimated market size (the proportion of potential adopters that had an initial positive individual assessment $I_i > 0$), and the y-axis plots the actual proportion of the population that adopted (those with a positive overall assessment, $Y_{ijt} > 0$, after diffusion completed,). As expected, the final extent of adoption for innovations with 0 novelty was perfectly predicted by the initial market size estimate (every potential adopter that had an initial positive assessment ultimately adopted, but no one else did).

Figure 3. Analysis of Simplified Baseline Simulation Model

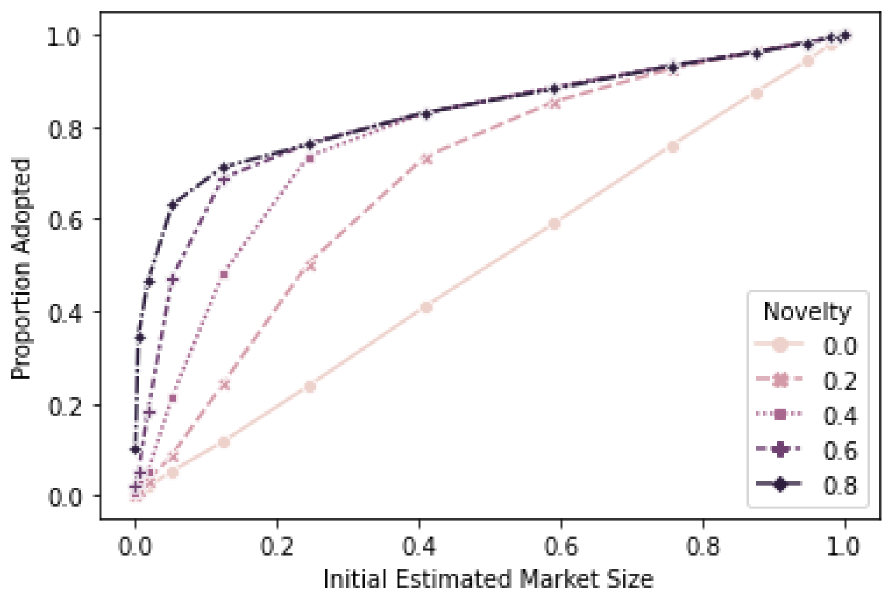


Figure 3 shows that high-novelty products can ultimately achieve the same level of adoption despite much lower initial estimated market sizes. In this version of the model, increasing novelty led to increased adoption, so that for any initial estimated market size, the highest market penetration was achieved by the most novel innovation. To take one dramatic example, consider that a 0 novelty innovation with an initial estimated market size of 0.5 ($\mu = 0$) would achieve the same 50% market penetration as a 0.6 novelty innovation with an estimated market size of less than 0.1 ($\mu = -0.15$).

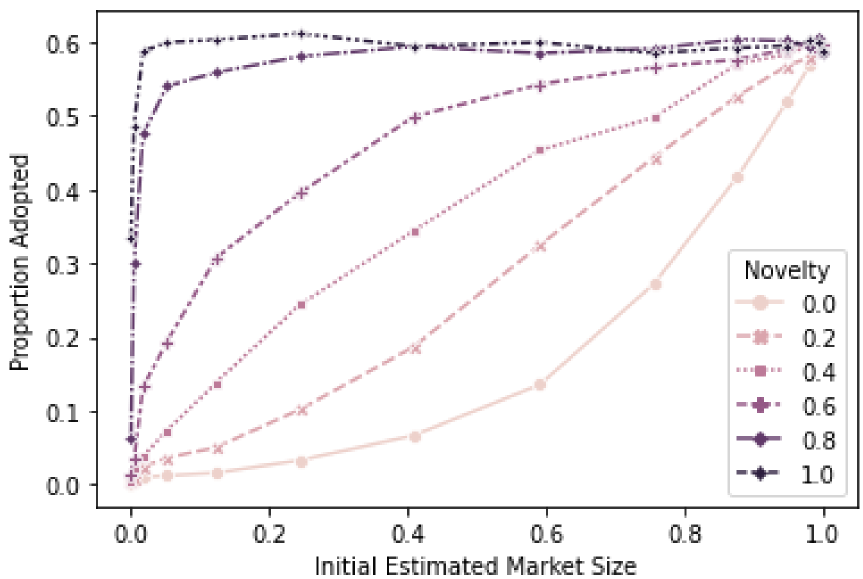
This version of the simulation serves as a useful baseline, but it makes strong assumptions that may not generalize across very many contexts. In particular, it assumes that each potential adopter has 1) perfect information about the existence of the product, and 2) all neighbor endorsements are positive. These conditions may apply to some dramatic examples, such as the opening example of Apple launching AirPods. Apple is prominent (close to perfect information), and has the uncanny ability to generate products that people love (most neighbor endorsements will be positive). But not every company is Apple, and not every product is AirPods. In the next sections, I relax these assumptions to make the model more generalizable.

Model 2: Main Simulation

Here, I present the main version of the simulation model. The simplified baseline model, presented above, assumed that all potential adopters were aware of the innovation at all times. But often, potential adopters have imperfect information—they are unaware of the innovation until others around them start using it. Here, I present a modified version of the simulation in which potential adopters are not aware of a product until at least one neighbor adopts. The simulation begins by randomly assigning adoption to one potential adopter that has a positive initial assessment.

Figure 4 plots the extent of final adoption for various combinations of the parameters using this version of the simulation. The imperfect information modification yields two meaningful changes relative to the results from the simplified baseline model depicted in Figure 3. First, imperfect information curtailed diffusion relative to the baseline model. In this modified model, the maximum proportion of final adopters had a lower ceiling at about 0.6, as compared to 1.0 in the baseline model (Figure 3). Therefore, on an absolute basis, most initial estimated market sizes exceeded the final extent of adoption (with the exception of highly ambiguous products with estimated market sizes less than 0.5).

Figure 4. Analysis of Main Simulation Model



Second, varying the novelty parameter N_j yielded even larger differences between the estimated adoption and the final extent of adoption than in the baseline model. High novelty innovations' initial market size estimates dramatically underestimated final adoption, whereas low novelty innovations' initial market sizes overestimated final adoption. This was especially true for products with estimated market sizes less than 0.5.

For an intuitive interpretation of these results, consider the illustrative case of a company that typically launches products that are 0.4 novelty. In this company, an accurate technique for estimating the final extent of diffusion could consist of two steps: 1) determine what proportion of potential adopters would adopt the new innovation according to their initial individual assessment; then 2) multiply that number by about 0.7 to get an accurate estimate of the extent of final diffusion.

For example, for a product that initially appeals to 40% of potential adopters, they would estimate about 30% final adoption (see the line corresponding to 0.4 novelty in Figure 4). That would be a fairly accurate forecast. But if the company applies this same technique to a more novel product (0.6 novelty), the 30% estimate would significantly underestimate the final extent of diffusion—it will actually diffuse to 45% of the potential adopters. And if it applies the same rule to a less novel product (0.2 novelty), the 30% estimate will overestimate the final extent of diffusion—it will only diffuse to about 18% of the potential adopters.

In summary, for products that are more novel than what the firm is used to, a relatively larger portion of potential demand does not exist until after the product diffuses, thereby downward biasing estimates of market demand for novel products. And on the other hand, for products that are less novel than what the firm is used to, customers are more certain of their evaluations, but the firm launching the new product may take for granted that the product will diffuse at the same rate as their typical product, and thus may make upward biased estimates of market demand. Taken together, these dual forces lead to a systematic inversion of market size expectations.

This main model is more generalizable than the baseline model, but still makes an assumption that may limit applicability to some contexts: that neighbors will all give positive endorsements for the innovation. This may hold for products that people love, but often the difficulty of innovation is that it is difficult to know *ex ante* which products people will love. In the next section, I will relax that assumption to generalize the model to cases where it is not known how much people will love the product once they adopt.

Model 3: Negative Endorsements Modification

In the prior versions of the model, all endorsements from adopting neighbors were positive. This may be a realistic assumption for products that generate positive experiences, but often it is difficult to tell *ex ante* which products will generate positive experiences. In this section, I modify the model so that it is possible for adopters to have negative experiences with the innovation. This version of the model will give a more realistic comparison of novel vs. non-novel products when it is not known *ex ante* that the innovation will generate positive endorsements.

In this version of the model, I keep the condition of imperfect information from the prior model, and I change the neighbor adoption to the proportion of positive endorsements minus the proportion of negative endorsements from neighbors (Abrahamson and Rosenkopf 1997):

$$P_{t-1} = \frac{Positive\ Neighbors_{t-1} - Negative\ Neighbors_{t-1}}{Total\ Neighbors}$$

In this version of the model, each neighbor that has adopted gives a positive endorsement if their post-adoption individual assessment, I_i^{post}, is >0, and a negative endorsement if I_i^{post}<0. Post-adoption individual assessment I_i^{post} for is drawn from a normal distribution centered on their initial individual assessment I_i, and with a standard deviation of N_j (more novel products vary more from the initial assessment):

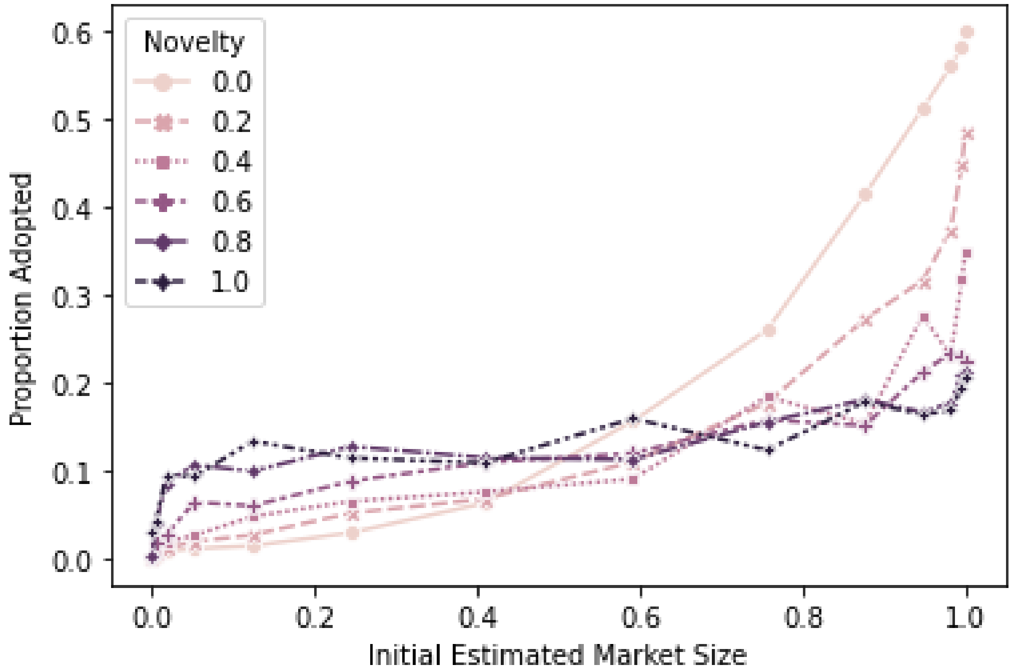
$$I_i^{post} \sim N(I_i, N_j)$$

For example, suppose there is a new innovation with novelty N_j=0.5, and potential adopter Alice has a neighbor, Bob. Alice has a negative initial assessment I_i =-0.05. Bob has a positive initial assessment I_i =0.05, so he adopts the new innovation. After he adopts, he realizes a post-adoption individual assessment of I_i^{post}=-0.1 (which was drawn from the normal distribution $N(0.05,0.5)$. Since I_i^{post}=-0.1<0, he has a negative post-adoption endorsement, which makes Alice even less likely to adopt.

Figure 5 plots the extent of final adoption for this version of the model. It shows that allowing for negative endorsements yields two meaningful changes relative to the results from the previous model. First, as expected, adding negative endorsements again curtailed diffusion relative to the prior model. In this model, maximum diffusion extent was still 0.6 for $N_j = 0$, but lower for higher values of N_j. So in most scenarios, initial estimated market sizes even more dramatically exceeded the final extent of adoption.

Second, in this version of the model, it was possible for low novelty products to perform better than high novelty products. In general, high novelty products still performed better when the initial proportion of positive assessments was less than 0.5, but low novelty products performed better when the initial proportion of positive assessments exceeded 0.5. However, it may be the case that this highest-performing scenario (a low novelty product with high estimated market size) will rarely occur in competitive markets. I elaborate on this possibility in the Discussion section.

Figure 5. Analysis of Modified Simulation (with Negative Endorsements)



In summary, unless a non-novel product is assessed positively by most potential adopters (which will be contested in competitive markets), the main takeaway of the model remains the same as before: high-novelty products can ultimately achieve the same level of adoption as low-novelty products despite much lower initial estimated market sizes. This result holds, even when allowing for negative experiences with the product.

Synopsis of Simulation Results

In summary, all versions of the model, comparing across hundreds of simulation scenarios, yield the same main takeaway: on average, high-novelty products can ultimately achieve the same level of adoption as low-novelty products, despite lower initial estimated market sizes. This is true even when allowing for imperfect information and negative endorsements from adopters (with the exception of non-novel innovations assessed positively by most potential adopters in Model 3).

Assuming that estimates of final diffusion are made based on some monotonic function of initial proportion of adopters with positive assessments, the main model and the modified model both yielded another takeaway: novel products overshoot initial expectations, while non-novel products undershoot initial expectations. Therefore, the model provides a pure demand-side explanation for the market size inversion puzzle that novel products overperform initial expectations while non-novel products underperform.

5. Empirical Validation

I empirically validate several predictions of the model using a dataset of sales diffusion for 1,600 Consumer-Packaged Goods (CPG) products from 2010-2016. First, I validate my model's assumption that customers rely more on neighbors' evaluations for novel products by demonstrating that novel products' diffusion curves more closely follow a socially-driven pattern. And second, by combining the product data with firm résumé data, I validate key implications of the model. Using the combined dataset, I show that firms that rely more on quantitative market sizing are more likely to fall into what I call the "market size trap": they are less likely to produce highly novel products, and less likely to produce breakthrough products, because these products appear relatively less attractive when relying on quantitative market sizing.

Data

I constructed a dataset that provides both product-level attributes and sales, as well as an in-depth view of the quantitative methods used in the innovation process within each organization. The final dataset is the result of an extensive data-collection process involving merges of several product-level and organization-level datasets, matched to aggregated employee-level information to track about 1,600 product launches at 41 large CPG organizations during the period 2010–2016. I give only a brief summary of the essential details of the dataset here, but extensive documentation for the source dataset is available in Chapter 1 of this dissertation (Allen 2022). This paper uses a subsample of products with higher quality data on product novelty and diffusion, as described below.

I constructed measures of product sales diffusion using the Nielsen Retail Measurement Services (RMS) scanner dataset.[24,25] To avoid counting minor changes to existing products as new products, I aggregated UPC barcode-level data up to the brand/product-module level.[26] I matched products to firms using the GS1 UPC-company-matching database.

I constructed measures of each organization's use of quantitative market sizing using résumé job description text from a popular online career networking website (DeSantola et al. 2020).[27] Because my aim was to measure quantitative analysis as it pertained to product innovation, I extracted résumé text only for employees whose reported job functions included "product management" or "research," using the networking website's internal job-function classification

[24] These data were accessed through the Kilts-Nielsen Data Center at the University of Chicago Booth School of Business.

[25] The Nielsen RMS data set is one of the most comprehensive and representative point-of-sale retail datasets available. It covers most sales from 40,000 distinct stores belonging to 90 retail chains across 371 Metropolitan Statistical Areas (Argente et al. 2018). The 2010–2016 sample includes, on average, $234 billion in annual sales, representing roughly 2% of annual U.S. household consumption.

[26] More specifically, I summed sales each quarter for all unique combinations of brand_descr and product_module_descr in the Nielsen data. Results are robust to UPC-level analyses.

[27] The website is arguably among the most representative data sources for the résumés of white-collar workers in the United States. As of 2018, it had over 150 million U.S. profiles.

system.[28] I obtained 101,919 unique employee résumés describing 182,403 positions[29] from 2010 to 2016 for the CPG firms in my sample.

To construct my primary dataset, I merged all of these data sources at the product level. For the test of the model implications, I also constructed a firm-level panel dataset. Both datasets include the same sample of 1,618 new products launched by 41 consumer product manufacturers (NAICS codes 31, 32, and 33) in the Nielsen data for 2010–2016, retaining only firm-years with sufficiently large samples of employee résumés (over 100 product/research-related résumés). Because meaningful text analysis required using only firms with large numbers of innovation-related résumés, the final sample of firms represents only relatively large and established CPG firms. About 80% of the firms in the sample belonged to the Fortune 1000 or Global 500 in 2016.

Measures

Dependent Variable(s). I measure the extent to which each product *i* diffused via a social process as the Q-coefficient of the Bass diffusion curve for each product (*Q-coefficient of Social* $Diffusion_i$). The Bass model is a differential equation that describes how new products are adopted in a population (Bass 1969). Two of the key parameters of the model are *q*, the coefficient of social diffusion, and *p*, the coefficient of external influence. The higher the level of *q*, the more adoption is driven by imitators, and the more the distribution of adoption resembles the traditional s-shaped

[28] The classification system uses profile information to classify individuals' job functions. Its classifications are not mutually exclusive (i.e., someone may be classified in both "product management" and "marketing"), which is important because different companies often assign different job titles to the same set of tasks (Baron and Bielby 1985, Bechky and Chung 2018, Sandholtz et al. 2019). My data-collection method imposed a cap of 1500 on the most relevant (according to the categorization algorithm) product- and research-related employees per firm; this restriction was only reached for one firm. In my sample, approximately 24% of the job titles of employees directly involved in the development and commercialization of products were explicitly labeled as research/R&D, 22% as product/innovation/strategy, 10% as marketing/brand, 9% as engineering, and the rest largely a mix of data analytics, IT, operations, and design. The networking website declined to provide further detail about its occupation-categorization algorithm; the company views the algorithm as a secret sauce for selling to HR professionals (its primary customers). The company claimed that the algorithm "has been refined over years and years of experimentation and research" by its teams of data scientists to make it as accurate as possible.

[29]There were more positions than unique employees because employees moved within and between firms.

adoption curve (mathematically, diffusion with high *q* and low *p* resembles a logistic distribution). The higher the level of *p*, the more adoption is driven by so-called "innovators"—those who purchase the product based on their own preferences or external influences like marketing, rather than social endorsement (mathematically, diffusion with high *p* and low *q* resembles an exponential distribution). I derive the *q* and *p* coefficients for each of the products in my dataset by fitting a Bass model to the first 8 quarters of sales volume. The average R^2 of the model in my final product sample was 0.68. To ensure a quality fit, I only used products that ever sold at least 100 units in the Nielsen data, and that did not return null values when fitting the Bass model. Together, these restrictions reduced the sample from 3,502 products to 2,436 products. In the firm-year panel dataset, this variable is aggregated by taking the mean *Q-coefficient of Social Diffusion* for all products launched by the firm each year.

Using the firm-year panel data, I also analyze dependent variables that capture commercial success. First, I define a *High Commercial Success$_i$* product as a one that achieved the top 5% of new product sales in their product group. Following prior work, new product sales is defined as the sum of dollar sales of product *i* in the two quarters ending the 2-year post-launch period (Allen 2022). I used sales in the last two quarters of the 2-year post launch period to avoid confounding signals from the initial size of the product launch, which is more likely to be endogenously determined by the firm than later sales (Bass 1969).

Finally, I define the variable *Breakthrough Innovation$_i$*, as a high commercial success product that also had some product novelty (*Product Novelty* >0; measure described below) and was characterized by diffusion process that was relatively socially driven (a Bass *q/p ratio* above the median) (Allen 2022). By including a requirement for some product novelty, this measure helps capture new products that are not merely exploiting prior successes. By restricting to products that relied relatively more on social diffusion the measure helps capture innovations that are not merely "replacement" products (Bass 1969, Hannigan et al. 2019).

Independent Variable(s). To validate the assumption of the model that more novel products are more likely to diffuse socially, I calculate *Product Novelty*$_i$ for product *i* as the percentage of product attributes (recorded by Nielsen) that were new-to-the-market (Allen 2022). Markets were defined by Nielsen product modules.[30] Due to incomplete product-attribute data, I was only able to calculate the novelty of a subset of the total number of products launched, which further reduced the sample from 2,436 new products to 1,618 new products.[31]

To test the implications of the model, I measure the extent to which each organization uses quantitative market research in the innovation process, calculated as the mean number of "quantitative" words in employee résumés' innovation-related sentences (for the list of "quantitative" words, see Appendix B). Although not a direct measure of initial market size expectations, the most common quantitative analyses such as surveys, market tests, concept tests, and forecasting are all methods that rely on potential customers' current assessments of a new product. Therefore, I argue that the extent of reliance on quantitative analysis in the innovation process is a reasonable proxy for reliance on customers' initial assessments. To construct this

[30] In the Nielsen data, product modules are more granular product category classifications than product groups. Product novelty is more comparable within modules, because products within a module share the same number of potential product attributes. Consider two illustrative examples from the heavy-detergents product module, which has 6 relevant attributes: product type, form (e.g., liquid, liquid pac), container (e.g., bag, bottle), type (e.g., with bleach, with stain removal), scent, and size. When a new product enters the heavy-detergents category with a never-before-seen scent and container, its index is calculated as ("# new attributes") /("# detergent attributes") = (2)/(6) = 0.33. If the brand consists of multiple UPCs, the measure is calculated as the mean *Product Novelty*$_i$ of all UPCs launched in the first 2 years of the product's life. In additional analyses, I also measured product novelty using the hedonic pricing adjustments used in Argente, Lee and Moreira (2018) and Granja and Moreira (2019), with nearly identical results (correlation between measures >0.95).

[31] To ensure the quality of the measure, I used only products with high-quality product-attribute data. I included only product modules that had at least 50 distinct UPCs with over $100 in sales, and (after standardizing NA values and merging attribute values with at least 0.9 Jaro-Winkler textual similarity) kept only attribute values that appeared at least 5 times and attributes with at least 5 distinct values. I then kept only individual UPC codes that had non-NA values for at least 1/3 of their recorded product attributes, and had at least 3 recorded product attributes.

measure, I filtered the résumé text to include sentences with at least one word related to a product innovation (see Appendix B). Overall, 151,727 of 588,301 sentences (about 25%) in my résumé sample included a product-innovation-related word. The measure *Quantitative Analysis*$_{ft}$ was calculated for each firm *f* in year *t* by taking the mean number of quantitative words within these product-innovation-related sentences. For additional details on this measure, see Allen 2022.

Control Variables. At the product level, I also control for the P-coefficient of External influence (the Bass model's parameter that captures direct marketing or advertising-based diffusion, rather than social influence) and the sales for each product in the first two quarters. I also control for several firm-year level factors, including the number of new products introduced by the firm that year, total annual sales (in $ millions), the total stock of products, and the number of résumé profiles at the firm that year. I lag each of these measures by one year, and take the natural log to normalize skewed distributions.

Table 1 presents summary statistics using product-level data. The mean Q-coefficient was 0.55, slightly higher than a typical Q-coefficient from other studies, but still within the expected range (Sultan et al. 1990). The mean P-coefficient was almost exactly the typical value around 0.034. The mean level of product novelty was 0.1, meaning on average, 10% of product features were new-to-the-market. About 26% of products were considered a "high commercial success", and 8% were breakthroughs. The mean level of quantitative analysis was 0.11, meaning the typical number of quantitative words in a résumé's innovation-related sentences was 0.11. The mean level of a new product's sales in the first 2 quarters was nearly a million, and companies had on average about 1000 innovation-related résumé profiles, 420 products, and almost $900 million in sales recorded in the Nielsen data (a representative sample of total sales).

Table 2 presents a correlation matrix using the same product-level data. Most variables have low levels of correlation, with the notable exception of the measures of company size (number of products and total sales. It should also be noted that, as expected, the First 2 quarters of sales are much more highly correlated with High Commercial Success (top 5% sales) than with a Breakthrough (top 5% sales with novelty and higher social diffusion).

Table 1. Summary Statistics

Statistic	N	Mean	St. Dev.	Min	Median	Max
Q-coefficient of Social Diffusion	1,618	0.551	0.426	0.002	0.490	13.553
Product Novelty	1,618	0.102	0.128	0	0.1	1
Breakthrough	1,618	0.080	0.271	0	0	1
Commercial Success	1,618	0.257	0.437	0	0	1
Quantitative Analysis	1,618	0.110	0.041	0.031	0.105	0.353
P-coefficient of External Influence	1,618	0.034	0.075	0.00002	0.025	2.432
First 2Q Sales (in $ thousands)	1,618	908.8	2,417.4	0.001	293.7	61,865.2
Company Profiles (in thousands)	1,618	1.115	0.715	0.187	1.022	2.603
Company Products	1,618	420.271	283.929	8	318.5	960
Company Sales (in $100 millions)	1,618	8.662	8.891	0.0002	3.987	29.610

Table 2. Correlations

	1	2	3	4	5	6	7	8	9	10
1 Q-coefficient of Social Diffusion	1									
2 Product Novelty	0.095	1								
3 Breakthrough	-0.032	0.060	1							
4 Commercial Success	-0.132	-0.074	0.500	1						
5 Quantitative Analysis	0.008	0.0002	0.051	0.111	1					
6 P-coefficient of External Influence	0.584	0.067	-0.075	-0.068	0.005	1				
7 First 2Q Sales	-0.085	-0.039	0.075	0.337	0.101	0.018	1			
8 Company Profiles	-0.030	-0.095	0.003	-0.012	-0.073	0.003	0.059	1		
9 Company Products	-0.016	-0.161	-0.081	-0.093	-0.156	-0.007	-0.035	0.616	1	
10 Company Sales	-0.001	-0.050	0.004	0.014	0.072	-0.008	0.056	0.829	0.649	1

Results

Table 3 presents regression results that validate the main assumption of the model: that novel products will more closely adhere to a social diffusion process. Column 1 displays the baseline result. For a given product, a 10% increase in product novelty is associated with about a

3.2% increase in the Q-coefficient of Social Diffusion. This result holds in column 2 when adding controls for the sales in the first 2 quarters, and for the P-coefficient of external influence, which proxy the direct marketing efforts for the product launch. The effect increases when using variables Winsorized at the 5% level, suggesting that the effect is not driven by outlier values (column 3). In column 4, the effect substantially increases when using a subsample of products with better fit parameters—specifically, products with Bass model R^2 values above the median of 0.73. And in column 5, the result holds when adding company controls, and company, year, and product group fixed effects. These results are consistent with the core assumption of the model, that a significant portion of growth for novel products will be realized only after spreading via word of mouth and social proof.

Table 3. Validating Model Assumption: Novelty and Social Diffusion

DV: Q-Coefficient of Social Diffusion	(1)	(2)	(3)	(4)	(5)
Product $Novelty_i$	0.316^{**}	0.373^{***}	0.434^{***}	0.598^{***}	0.344^{*}
	(0.115)	(0.110)	(0.099)	(0.134)	(0.136)
First 2Q $Sales_i$		-0.020^{***}	-0.019^{***}	-0.031^{***}	-0.022^{**}
		(0.005)	(0.004)	(0.006)	(0.008)
P-coefficient of External $Influence_i$		-0.124^{***}	-0.141^{***}	-0.026^{+}	-0.140^{***}
		(0.012)	(0.013)	(0.014)	(0.017)
Company $sales_{ft-1}$ (in $100 millions)					0.157
					(0.122)
Company $Products_{ft-1}$					0.174^{+}
					(0.093)
(Intercept)	-0.747^{***}	-0.988^{***}	-1.061^{***}	-0.383^{***}	
	(0.016)	(0.087)	(0.081)	(0.103)	
Company, Year, Product Group FE					YES
Subsample: Bass R^2 Above Median				YES	
Winsorized at 5%			YES		
N	1618	1618	1618	1233	1618
Adj. R^2	0.004	0.089	0.113	0.085	0.129
Within R^2					0.113

All variables are logged to adjust for skew and normalize the residual errors. Robust standard errors reported in parentheses. $^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$; $^{+}p < 0.1$

I also validate the implications of the model. According to the model, relatively novel products will appear relatively less attractive according to *ex ante* quantitative estimates of market size based on customers' initial individual assessments. Therefore, I suspect that firms that rely more on quantitative analysis will be less likely to launch novel products, and less likely to launch breakthroughs—a result I call the "market size trap". Table 4 confirms these implications. Although Column 1 shows an insignificant relationship between quantitative analysis and mean product novelty, Column 2 does confirm a significant negative relationship between quantitative analysis and the rate of highly novel (top quartile) products launched by a firm. A 10% increase in quantitative analysis was associated with a 8.8% decrease in the rate of highly novel products launched. And although Column 3 shows a (not quite significant) positive relationship between quantitative analysis and the rate of commercial successes (top 5% sales among all Nielsen products in the same product category), there is a significant negative relationship between quantitative analysis and the rate of breakthroughs (products with top 5% sales that also had some novel features and relied relatively more on social diffusion). A 10% increase in quantitative analysis was associated with a 2.5% decrease in the rate of breakthrough products.

Table 4. Model Implication: The Market Size Trap

	Novelty		Product Success	
	DV: Mean Product Novelty	**DV: Highly Novel Rate**	**DV: High Commercial Success Rate**	**DV: Novel Breakthrough Rate**
	(1)	**(2)**	**(3)**	**(4)**
Quantitative Analysis$_{ft-1}$	-0.029	-0.877*	0.639^{+}	-0.250*
	(0.107)	(0.341)	(0.327)	(0.127)
Company Profiles$_{ft-1}$	0.006	-0.068**	-0.002	0.012^{+}
	(0.007)	(0.025)	(0.017)	(0.007)
Company Sales$_{ft-1}$	-0.001	0.002	0.031***	0.006^{+}
	(0.004)	(0.009)	(0.009)	(0.003)
Company Products$_{ft-1}$	0.001	0.032	-0.045*	-0.009
	(0.006)	(0.021)	(0.023)	(0.010)
Industry FE	YES	YES	YES	YES
Year FE	YES	YES	YES	YES
N	232	293	293	293
Adj. R^2	0.106	0.075	0.093	0.036

All variables are logged to adjust for skew and normalize the residual errors. Robust standard errors reported in parentheses. $^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$; $^{+}p < 0.1$

6. Discussion

This paper advances a theory for the "market size inversion" puzzle that novel products overperform and non-novel products underperform initial expectations. I proposed a model in which customers' adoption decisions depend relatively more on neighbors' endorsements for ambiguous products, so a larger portion of estimated demand does not exist until after the diffusion process. Simulating the model across many scenarios, I showed that high-novelty products can ultimately achieve the same level of adoption as low-novelty products, despite lower initial assessments from potential adopters. Therefore, assuming market size estimates are based on potential adopters' initial assessment of the new product, estimates for novel products (which are more ambiguous to evaluate) will be biased downward.

I empirically validated the model using sales and attribute data from about 1,600 consumer product launches, combined with measures firms' use of quantitative market sizing techniques derived from résumé data. First, I validate that product novelty predicts the extent to which products diffuse socially, a key assumption of the model. Second, I validate the implication of the model which I call the "market size trap": firms that rely more on quantitative analysis in the innovation process (proxying market size estimates based on customers' initial assessments of a new product) launch lower rates of highly novel products, and lower rates of breakthrough products. These findings contribute new theoretical insight to two streams of strategy research.

Complementing the Competition-based theory of Market Size Inversion

A primary contribution of the paper is a demand-side model which explains the market size inversion puzzle (i.e., novel products overperform initial expectations while non-novel products underperform), purely as a function of diffusion dynamics. My diffusion-based model could be viewed as an alternative explanation for the prevailing competition-based explanation for market size inversion: that demand uncertainty (about features, uses, regulation, price points, etc.) is a competitive barrier to entry, which provides valuable opportunities for uncontested innovative rents for those who have superior cognitive ability to perceive these opportunities (Denrell et al.

2003, Felin and Zenger 2017, Gavetti 2012). I show that this pattern arises even without explicitly accounting for competition, simply because customers' adoption decisions depend relatively more on neighbors' endorsements for novel products, so a larger portion of demand does not exist until after the diffusion process.

My diffusion-based model improves upon the prior model by further illuminating which products will be more likely to succeed *a priori*, and for understanding when market size estimates are useful. In my main model (Model 2), I showed that for a firm consistently launching products that are the same level of novelty, they can use initial demand to consistently estimate the final extent of adoption. But if they launch a higher novelty product, the same estimation technique will likely underestimate adoption by not sufficiently accounting for the demand that is realized post-diffusion; if they launch a lower novelty product, they will likely overestimate adoption by incorrectly assuming that the all of the interested potential adopters will become aware of the product with a less socially-driven diffusion process. Therefore, market size estimates based on demand prior to diffusion are only useful when comparing between products with similar levels of novelty. Because market size estimates are only useful when comparing between products with similar levels of novelty, firms that evaluate products with varying levels of novelty may underestimate the value of the novel products. Indiscriminately choosing the projects with the largest projected market sizes without regard to the comparability of the projections across products of varying novelty leads to the "market size trap" of launching fewer novel products that could become breakthroughs. But explicitly accounting for launching products with different levels of uncertainty could theoretically remove much of this uncertainty.

However, my diffusion-based model is most useful as a complement, rather than a substitute, for the prior competition-based model. Although my model can technically explain market size inversion alone, it is more likely that the forces of diffusion and competition are both at play in most situations. Furthermore, my model alone has difficulty explaining why in model 3 (perhaps the most generalizable model which includes the possibility of negative endorsements) firms will not always attempt to launch products with a low novelty product and high estimated market size (which had the greatest extent of diffusion). The competition-based theory explains why this is not possible. In competitive markets, unless there is some other barrier to entry (such as a competitive technology or brand), such unambiguously desired products will be readily apparent to competitors, and therefore highly contested. Adding competition to the model in accordance with theories of competitive strategy would significantly curtail diffusion for the less ambiguous products with high estimated market sizes. Formally adding competition to the diffusion-based model is one potential extension of this model and an avenue for future research.

My model also complements the competition-based theory by further refining our understanding of who will successfully innovate. Prior work suggests that those with foresight that is superior to competitors will capture innovative gains (Felin and Zenger 2017, Gavetti 2012). My model further specifies that successful innovators have foresight that is superior to customers, and are able to perceive which products will have high demand post-diffusion, despite low initial expectations. This requires insight beyond even a perfect understanding of customers' own demand, for which there is great uncertainty about their own preferences prior to the launch of a product. In essence, the model is a concrete formalization and explanation of the principle behind the popular (albeit apocryphal) Henry Ford quote that "if [he] had asked people what they wanted, they would have said faster horses" (Vlaskovits 2011).

Social Diffusion as a Source of Uncertainty for Novel Products and Nascent Technologies

The paper also makes a secondary contribution to research on the sources of market uncertainty for novel products and nascent technologies. This stream of research has identified demand uncertainty for a new technology from a variety of sources: from customers who have not tangibly engaged with the new product and so are not aware of how the new features will benefit them (von Hippel 1986); from new products "framing" that may or may not be consistent with prior understanding (Anthony et al. 2016, Benner and Tripsas 2012, Bingham and Kahl 2013); from a lack of awareness of the future uses of the technology which have not been identified yet (Bresnahan 2010, Rosenberg 2009); from evolving trajectories of functional needs and price points evolve (Adner and Levinthal 2001, Adner and Snow 2010, Adner and Zemsky 2006, Christensen and Bower 1996); and from regulatory and institutional dimensions make it unclear whether the new product is legitimate or safe (Gao and McDonald 2022, Ozcan and Gurses 2018).

Thus, prior research has inherently taken a technological (e.g., technical features, use set) or market oriented (e.g., price, regulatory) perspective to understanding customer adoption decision. My paper illuminates additional uncertainty that arises because adoption (especially of ambiguous products) can be a highly *social* process. Even if other sources of uncertainty are removed (i.e., customers have tangible experience with a product, the product meaning is framed appropriately, not accounting for unknown uses of the technology, even without changing functional needs or regulatory constraints), a significant amount of uncertainty exists purely because of the social diffusion process. A key insight of the model is that demand for novel products does not even exist until it is created by neighbors' adoption, which only then resolves other potential adopters' uncertainty about the new product. So even with a 100% accurate, unbiased survey of customer demand, and no other sources of uncertainty, the diffusion dynamics for products with varying levels of novelty create additional uncertainty.

Therefore, without accounting for this diffusion dynamic, there is no purely empirical market research technique that can result in a more accurate estimate of final market size. However, it is possible that a model that explicitly accounts for the novelty or ambiguity of a product, and thus accounts for the extent to which it will socially diffuse, could resolve a significant amount of demand uncertainty prior to launch. Further developing techniques for incorporating the novelty/ambiguity of a product into product forecasts could be an avenue for future research.

7. Conclusion

Despite millions of dollars invested into market research, innovating firms often miss potential breakthroughs and launch failed products. One partial explanation for this phenomenon is that they may insufficiently account for how novel products diffuse differently. Novel products provide opportunities for largest innovative rents, but because they rely more on social endorsement than other products, they appear less attractive prior to the diffusion process. Without sufficiently accounting for this difference, firms may underestimate the potential of novel breakthroughs, while overestimating the potential of supposedly safe products.

References

Abrahamson E, Rosenkopf L (1997) Social Network Effects on the Extent of Innovation Diffusion : A Computer Simulation. *Organ. Sci.* 8(3):289–309.

Adner R, Levinthal D (2001) Demand heterogeneity and technology evolution: Implications for product and process innovation. *Manage. Sci.* 47(5):611–628.

Adner R, Snow D (2010) Old technology responses to new technology threats: Demand heterogeneity and technology retreats. *Ind. Corp. Chang.* 19(5):1655–1675.

Adner R, Zemsky P (2006) A demand-based perspective on sustainable competitive advantage. *Strateg. Manag. J.* 27(3):215–239.

Albert R, Barabási AL (2002) Statistical mechanics of complex networks. *Rev. Mod. Phys.* 74(1):47.

Allen RT (2022) *Methodological Pluralism and Innovation in Data-driven Organizational Cultures*

Anthony C, Nelson AJ, Tripsas M (2016) "Who are you?… I really wanna know": Product meaning and competitive positioning in the nascent synthesizer industry. *Strateg. Sci.* 1(3):163–183.

Argente D, Lee M, Moreira S (2018) Innovation and product reallocation in the great recession. *J. Monet. Econ.* 93:1–20.

Banerjee A, Chandrasekhar AG, Duflo E, Jackson MO (2013) The diffusion of microfinance. *Science (80-.).* 341(6144):1236498.

Baron JN, Bielby W (1985) Organizational barriers to gender equality. *Gend. life course*.

Bass FM (1969) A new product growth for model consumer durables. *Manage. Sci.* 15(5):215–227.

Bechky BA, Chung DE (2018) Latitude or latent control? How occupational embeddedness and control shape emergent coordination. *Adm. Sci. Q.* 63(3):607–636.

Benner MJ, Tripsas M (2012) The influence of prior industry affiliation on framing in nascent industries: The evolution of digital cameras. *Strateg. Manag. J.* 33(3):277–302.

Bingham CB, Kahl SJ (2013) The process of schema emergence: Assimilation, deconstruction, unitization and the plurality of analogies. *Acad. Manag. J.* 56(1):14–34.

Bresnahan T (2010) General purpose technologies. *Handb. Econ. Innov.* 2:761–791.

Christensen CM, Bower JL (1996) Customer Power, Strategic Investment, and the Failure of Leading Firms. *Strateg. Manag. J.* 17(3):197–218.

Delre SA, Jager W, Janssen MA (2007) Diffusion dynamics in small-world networks with heterogeneous consumers. *Comput. Math. Organ. Theory* 13(2):185–202.

Denrell J, Fang C, Winter SG (2003) The economics of strategic opportunity. *Strateg. Manag. J.* 24(10 SPEC ISS.):977–990.

DeSantola A, Gulati R, Zhelyazkov PI (2020) External Interfaces and Internal Processes: Market Positioning and Divergent Professionalization Paths in Young Ventures.

Felin T, Gambardella A, Stern S, Zenger T (2019) Lean startup and the business model: Experimentation revisited. *Forthcom. Long Range Plan. (Open Access)*.

Felin T, Zenger T (2017) The Theory-Based View: Economic Actors as Theorists. *Strateg. Sci.* 2(4):258–271.

Felin T, Zenger T (2018) What sets breakthrough strategies apart. *MIT Sloan Manag. Rev.* 59(2):86–88.

Fingas R (2018) Apple sold 35 million AirPods in 2018, currently most popular "hearable"

brand. *Apple Insid.*
Gao C, McDonald R (2022) Shaping nascent industries: Innovation strategy and regulatory uncertainty in personal genomics. *Adm. Sci. Q.*
Gavetti (2012) Toward a behavioral theory of strategy. *Organ. Sci.* 23(1):267–285.
Geroski PA (2000) Models of technology diffusion. *Res. Policy* 29(4–5):603–625.
Granja J, Moreira S (2019) Product Innovation and Credit Market Disruptions. *SSRN Electron. J.*
Hannigan T, Haans RFJ, Vakili K, Tchalian H, Glaser VL, Wang M, Kaplan S, Jennings PD (2019) Topic Modeling in Management Research. *Acad. Manag. Ann.* 13(2):586–632.
von Hippel E (1986) Lead Users: a Source of Novel Product Concepts. *Manage. Sci.* 32(7):791–805.
Jin C, Song C, Bjelland J, Canright G, Wang D (2019) Emergence of scaling in complex substitutive systems. *Nat. Hum. Behav.* 3(8):837–846.
Kelly SM (2016) Apple AirPods review: Do they actually stay in your ears? *CNN Money*.
Miller C (2019) Tim Cook calls AirPods a 'cultural phenomenon,' Apple working hard to keep up with demand. *9to5Mac*.
Moore GA (1991) *Crossing the chasm*
Ozcan P, Gurses K (2018) Playing cat and mouse: Contests over regulatory categorization of dietary supplements in the United States. *Acad. Manag. J.* 61(5):1789–1820.
Patel N (2016) Apple AirPods hands (and ears) on. *The Verge*.
Rogers EM (1962) *Diffusion of innovations*
Rosenberg N (2009) Uncertainty and technological change. *Econ. impact Knowl.* (Routledge), 17–34.
Sandholtz K, Chung D, Waisberg I (2019) The double-edged sword of jurisdictional entrenchment: explaining human resources professionals' failed strategic repositioning. *Organ. Sci.* 30(6):1349–1367.
Schneider J, Hall J (2011) Why most product launches fail.
Sultan F, Farley JU, Lehmann DR (1990) A meta-analysis of applications of diffusion models. *J. Mark. Res.* 27(1):70–77.
Valente TW, Rogers EM (1995) The origins and development of the diffusion of innovations paradigm as an example of scientific growth. *Sci. Commun.* 16(3):242–273.
Vlaskovits P (2011) Henry Ford, innovation, and that "faster horse" quote. *Harv. Bus. Rev.* 29(08):2011.
Wang D (2019) How New Versions of Products Spread Differently Than Entirely New Products How New Versions of Products Spread Differently Than Entirely New Products.
Zellweger TM, Zenger TR (2021) Entrepreneurs as scientists: A pragmatist approach to producing value out of uncertainty. *Acad. Manag. Rev.* (October).

Chapter 3. The Limits of Experimentation for Product Innovation in Homogenous User Communities

1. Introduction

Recent research on product innovation highlights benefits for firms working with *user communities*, or loose collections of individuals who work voluntarily, collaboratively, and with little oversight to provide feedback on and share ideas about a domain of common interest (usually a product or a technology) (Shah, 2005; Shah & Nagle, 2020; Shah & Tripsas, 2007).[32] User communities are characterized by open membership, self-selection into tasks, collaboration, and self-generated rewards (Lakhani, 2016; Mollick, 2016; O'Mahony, 2003; Shah, 2006), with members engaging in such activities as socializing, help-seeking, help-giving, and peer production (Lakhani & Von Hippel, 2003; Shah & Tripsas, 2007). Firms that engage with these communities can source novel ideas from beyond their organizational boundaries to enhance innovation (Eklund & Kapoor, 2022; Piezunka & Dahlander, 2015). Shah and Nagle (2020) argue that "engines of innovative activity," which have historically resided within traditional firms and research institutions, increasingly occupy the community–firm nexus, or the interstitial spaces where organizations and user groups work in tandem to create something novel (p. 308). Collectively, researchers have framed engagement with user communities as an underappreciated resource or pathway that firms can utilize to improve innovation performance.

[32] Definitions vary; ours captures the key features that user communities share (see Shah and Nagle, 2020) and is keeping with prevailing popular and managerial usage (Lessig 2006; Antorini, Muñiz, and Askildsen 2012; Kost 2019). Unlike participants in crowd contests (Afuah & Tucci, 2012; Piezunka & Dahlander, 2015), users in communities typically collaborate, rather than compete, and are more apt to be intrinsically motivated to contribute (Boudreau & Lakhani, 2015; Lakhani, 2016).

This largely positive perspective on user communities prevails in research on business experimentation, which conceptualizes experimentation as a learning process involving deliberate variation of activities to resolve uncertainty (Miner et al., 2001; Thomke, 2003). Researchers have shown that successful innovations tend to arise within experimentation pathways enabled by communities: from the pool of users who provide individual feedback on products under development (Baldwin et al., 2006; Katila et al., 2017; Thomke & Von Hippel, 2002), and from the knowledge that flows from collective interaction among users (Bayus, 2013; Franke & Shah, 2003; Hienerth et al., 2014; Von Krogh et al., 2003). In short, a key premise of existing research is that new product experimentation in conjunction with communities will "revolutionize innovation" (Harhoff & Lakhani, 2016).

However, the literature has given less attention to the fact that many user communities are not representative of other users outside the community. Members self-select into the user community precisely because they are passionate about the community's focal product. Although this intrinsic motivation drives engagement in innovations catered to other like-minded people in the community, it is unclear whether products that appeal to such niche users will achieve commercial success with outsiders. On one hand, from a practitioner standpoint, entrepreneurs are encouraged to develop for an early passionate group of users, with admonitions from advisors to "build something 100 people love, not something 1 million people kind of like" (Shontell, 2013). On the other hand, some research in entrepreneurship and marketing suggests that early users tend to differ from those of mainstream consumers, and thus that catering too heavily to their preferences can diminish mainstream commercial appeal (Cao et al., 2021; Moore, 1991). These findings cast doubt on the assumption that soliciting feedback from user communities is unambiguously helpful for creating innovations that succeed in the market. Specifically, feedback from early users with unrepresentative preferences may persuade firms to sideline or abandon otherwise-promising ideas. These observations hint at a phenomenon overlooked by the literature on user communities: potential drawbacks of soliciting feedback from user communities characterized by a high level of what we call *community homogeneity*—the degree to which the community is characterized by a narrow, unrepresentative segment of the market.

We suggest that reaping the benefits of experimentation in new product innovation is contingent on the degree to which the community represents demand in the broader market. As a baseline assumption, we propose that user-community feedback has a positive main effect (on average) on a new product's growth in the market. But we hypothesize that community homogeneity negatively moderates this effect: responding to feedback from homogenous communities can actually dampen market growth (thereby reversing the main effect).

We explore our hypothesis in the context of PC video-game development, a highly uncertain economic domain due to unclear market preferences and product complexity (Bresnahan et al., 2014; Katila et al., 2022; Mollick, 2012; Yin et al., 2014; Zhao et al., 2018). We study games available from Steam, an online digital-distribution platform for PC games that facilitates over 60% of global game-industry sales (Edwards, 2013; Hruska, 2018). Steam's experimentation-promoting tool, Early Access, enables game manufacturers to build user communities around specific games before finalizing their underlying code; manufacturers can also test and modify early-stage product concepts.

We constructed a monthly panel dataset of nearly 400 Steam Early Access games that were under experimental development. Each Early Access game forms its own community, whose members engage in collective activities, as socializing, help-giving, help-seeking, and discussion (Lakhani & Von Hippel, 2003; Shah & Tripsas, 2007), that can inform subsequent game development and growth. Substantial within-game variation over time allows for detailed assessment of how user communities' features influence game adoption throughout the experimental development phase. To enrich and provide context for our findings, we supplemented this quantitative data by conducting 29 interviews with game developers, industry executives, and community members. We provide evidence in support of our baseline assumption: community feedback improves games' market growth on average. We also confirm our hypothesis that the effect is strongly negatively moderated by community homogeneity. These results support our theory that homogenous communities generate signals of market demand that, when incorporated into a product, restrict a game's appeal to audiences outside the community.

Our study extends theoretical understanding of the role of user communities in experimentation and feedback within new-product development. Our main contribution establishes a boundary condition on when firms are likely to realize benefits from experimentation with user communities. Prior research has examined how the characteristics of *firms* (such as communication channels) and those of *users*, such as network centrality (Dahlander & Frederiksen, 2012) and tenure in the community (Anthony et al., 2009), determine whether firms gain insights from user feedback (Foss et al., 2011). We provide a complementary perspective on how a *community* characteristic—homogeneity—moderates the experimentation/innovation relationship. Specifically, we theorize and show that more-representative communities magnify the benefits of experimentation (Felin & Zenger, 2014; Shah & Tripsas, 2007), but that ideas and solutions voluntarily offered by a well-meaning but homogenous community of users can derail the innovation's trajectory.

2. Theory and Hypothesis

2.1 How firms work with users and user communities

Several streams of research examine the role of users and user communities in innovation. One such stream explores *how* firms work with individual users and user communities. Collaboration takes several forms, including soliciting ideas from users (Bayus, 2013; Dahlander & Piezunka, 2014; Lilien et al., 2002), sourcing complementary innovation (Dahlander & Gann, 2010; Jeppesen & Molin, 2003), and contributing to community-based innovation (Dahlander & Magnusson, 2005; Dahlander & Wallin, 2006; Nagle, 2018). Lilien et al. (2002) illustrate the first approach in their study of 3M's "lead user process" for idea generation: the firm collected information from users about their needs and suggestions to guide internal innovation efforts. By contrast, Jeppesen and Mollin (2003) explore sourcing innovation directly from users by studying user-to-user and user-product interactions while playing a newly released video game. Using tools provided by the focal firm, users modified the game and produced new content that other users could then play. Users within the community thus facilitated a form of learning (on the part of the firm) in which tools and products were altered in novel ways.

2.2 Benefits of working with users and user communities

A second stream of literature explores the *benefits to firms* of working with individual users and user communities. A key finding is that firms benefit from the novelty of user-generated ideas (Poetz & Schreier, 2012; Schweisfurth, 2017). Schweisfurth's (2017) survey-based study of product development in the home-appliance industry finds that the product ideas of users external to the firm were rated by judges as higher-quality (more novel, with greater commercial potential) than ideas generated by users within the firm. Research on crowdsourcing has reached a similar conclusion (Afuah & Tucci, 2012; Lifshitz-Assaf, 2018; Piezunka & Dahlander, 2015). Lifshitz-

Assaf's (2018) study of NASA's crowdsourcing efforts found that a crowd solved several complex problems far faster than NASA scientists had anticipated. In one instance, a retired communications engineer from New Hampshire "stunned" NASA with a working method of predicting solar-particle events. More broadly, research suggests that firms themselves are more innovative when working with users (Chatterji & Fabrizio, 2012, 2014; Foss et al., 2011; Katila et al., 2017). Chatterji and Fabrizio's (2014) longitudinal analysis of 128 medical-device firms finds that those that collaborate with physicians (that is, users) during the innovation process are more likely to win FDA approval for new devices (see Smith and Shah, 2013, for a similar finding in the corporate-venture-capital industry).

Scholars have also begun to explore the benefits to firms of working with *communities* of users (Nagle, 2018, 2019). A key insight is that communities facilitate knowledge sharing, which in turn improves innovation outcomes. Shah and Tripsas (2007), in a study of entrepreneurship in the juvenile-products industry, find that the ventures that succeeded had engaged in collective creative activity (such as by socializing with each other, discussing shared needs, and building on each other's ideas) that had significantly enhanced the quality of their ideas. In a more recent empirical study, Nagle (2018) analyzes longitudinal data on 58 matched firms that work with user-community-developed open-source software: firms that contribute code to the user community benefit more from using the software than do non-contributors, suggesting that firms learn by collaborating with communities. Baldwin and Von Hippel's (2011) mathematical model shows an upper limit on the level of design costs that firms can incur (because of their need to profit); user communities, by contrast, can spread the cost of design across thousands of users.

Other scholars have begun to specify boundary conditions on when user communities will excel in innovation (Bremner & Eisenhardt, 2021; Felin & Zenger, 2014). For example, Bremner and Eisenhardt's (2021) case study of innovation in the nascent drone industry finds that user communities struggle to pursue innovation in contexts characterized by significant complexity, where firms have a relative advantage. Nevertheless, the prevailing perspective on user community innovation remains largely positive.

2.3 User communities, experimentation, and the commercial success of innovations

Despite this ample literature on the ideation benefits of engaging with user communities, questions remain about whether these benefits translate to commercially successful new products. In particular, the literature has given less attention to the fact that many user communities are often not representative of potential users outside the community. Members self-select into the user community based on a particular set of preferences. Although this intrinsic motivation drives engagement in the community, it is unclear whether products that appeal to such niche users will achieve commercial success with outsiders.

Findings from other literatures cast doubt on the assumption that soliciting feedback from user communities is unambiguously helpful for creating innovations that succeed in the market. Research on business experimentation posits that realizing the innovation-related benefits of experimentation may be contingent on the characteristics of users—and, by extension, on the particular community of users that a firm engages with (Camuffo et al., 2019; Felin et al., 2019; Felin & Zenger, 2014). And several anecdotes from other literatures highlight that adapting products in response to early adopters' preferences can decrease appeal in the target market (Cao et al., 2021; Moore, 1991; Rogers, 1962). In pursuit of a more comprehensive portrait, we offer formal hypotheses on how community homogeneity moderates the relationship between firms' responsiveness to user-community feedback and subsequent product performance.

2.4 Experimentation within user communities and commercial performance

We conceptualize experimentation in user communities as a process of iteration between users' feedback and product adaptation. It will be useful here to distinguish between an experiment and experimentation (see Thomke, 2003). *An experiment* is usually a specific one-shot event with clearly isolated treatment and control groups, like an A/B test. Two alternatives are compared; whichever alternative performs better is then implemented. This kind of experiment is relatively rare in user communities. Instead, product innovation in user communities entails *experimentation*—a dynamic, ongoing process in which firms try out new initiatives and respond to community feedback. Innovators formulate a plan to improve the product in keeping with community feedback, and then adapt it accordingly. They observe the community's reaction and decide how to respond.

Experimentation helps firms improve innovation outcomes by resolving *market uncertainty*, or lack of information about the preferences of potential customers (Toh & Kim, 2013). Market uncertainty hinders firms from determining how to design products (e.g., which features to include) to appeal to a sufficiently large audience at a sufficiently high price point (MacCormack et al., 2001). Experimentation can help fill this informational void by providing information about customers' preferences (Camuffo et al., 2019; F. Murray & Tripsas, 2004; Ozcan & Eisenhardt, 2009), which firms can use to update their beliefs about the market over time (Kapoor & Wilde, 2022). Firms can use a variety of methods, but feedback from users interacting with prototypes is especially likely to provide rich feedback that reflects demand reasonably well (Chatterji & Fabrizio, 2014).

Firms that experiment in collaboration with *communities* of users are likely to improve their commercial performance further. Communities provide users a forum on which to share their experiences with the focal product (Jeppesen & Molin, 2003; Shah & Tripsas, 2007). These interactions in turn enhance users' understanding of the product beyond what they could learn on their own (Autio et al., 2013; Jeppesen & Molin, 2003). Such understanding in turn equips users to provide firms better feedback and ideas. For example, Boudreau and Lakhani (2015) conduct a field experiment using a problem-solving challenge to examine whether competitors fare better when allowed to share information with each other. The authors find that competitors with access to information from others perform better even while exerting less effort. In addition to soliciting direct feedback (e.g., Lilien et al. 2002), firms can learn from social interactions between users. Such learning may provide more accurate insights than does direct questioning, because participants are less constrained. More broadly, learning via multiple channels, such as feedback, observation, and discussion, is more likely to generate insight (Bingham & Davis, 2012; Brown & Eisenhardt, 1997; McDonald & Eisenhardt, 2020).

In sum, prior work indicates that experimentation can reduce market uncertainty when the information it provides is incorporated into new products (MacCormack et al., 2001; Thomke, 2003; Yin et al., 2014). Experimentation with user communities is apt to multiply those benefits via the iterative interplay between feedback and firms' responsive product adaptation (Shah & Tripsas, 2007). Thus, our model's baseline premise is derived from prior research:

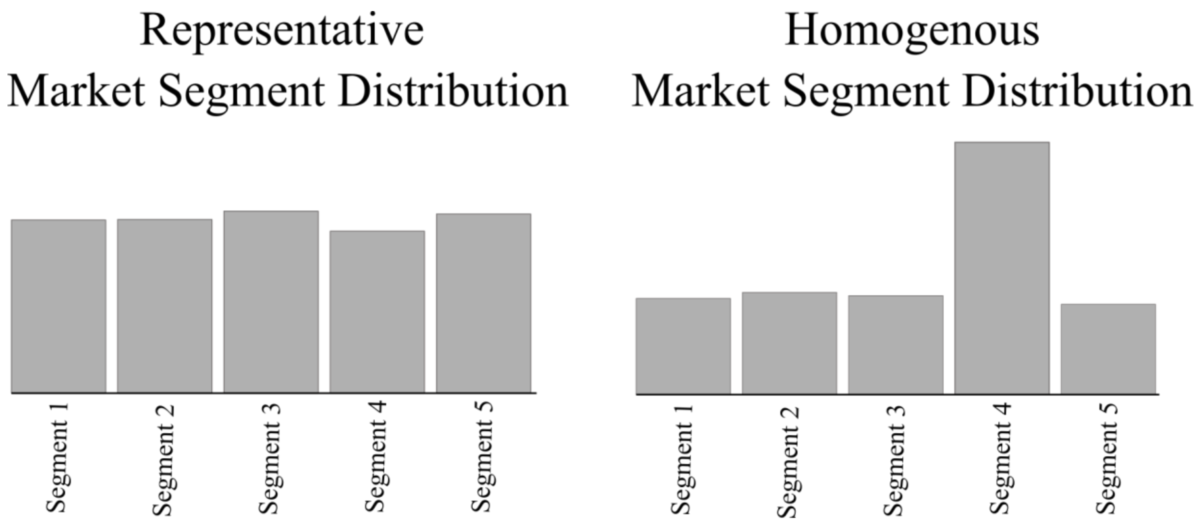
> *Baseline premise: On average, firms' responsiveness to community feedback increases new product revenue.*

2.4.1 The moderating role of community homogeneity

However, experimentation within homogenous communities may yield results that differ from the baseline established by the prior literature. Prior work on user communities has largely ignored the insight that early users of new products can differ significantly from outside mainstream customers. In work on marketing new high-technology products, for instance, Moore (1991) posited a chasm between the needs of early adopters and those of the mainstream market. The expectations of a new product's earliest users (so-called technology enthusiasts or visionaries) tend to differ strikingly from those of the early majority (or pragmatists); by implication, feedback from a group of ardent users could misdirect an innovation or mislead firms aiming for broader market acceptance (also see Franz, 1999). In a more recent study, Cao, Koning, and Nanda (2021) examined the online platform Product Hunt, where founders, developers, and designers provide feedback on other startups' early product ideas. The authors found that randomly oversampling feedback from males decreased female-focused products' chances of success; this finding implies that the characteristics of early users can distort views of new products' potential (also see Shah and Nagle, 2020). These observations also suggest that biased feedback from early users may persuade firms to sideline or abandon otherwise promising innovations, a potential shortcoming of soliciting feedback in user communities characterized by a high level of what we call *community homogeneity*.

This paper defines community homogeneity as the degree to which members of a community belong to market segments that are non-representative of the broader market. Figure 1 illustratively compares a homogenous distribution to a representative distribution of users' market segments. For the sake of parsimony, we conceptualize each user as belonging to a single market segment. As the figure shows, in a homogenous community, a disproportionate number of members belong to one segment of the market (or a small set of market segments).

Figure 1: Representative vs. homogenous communities

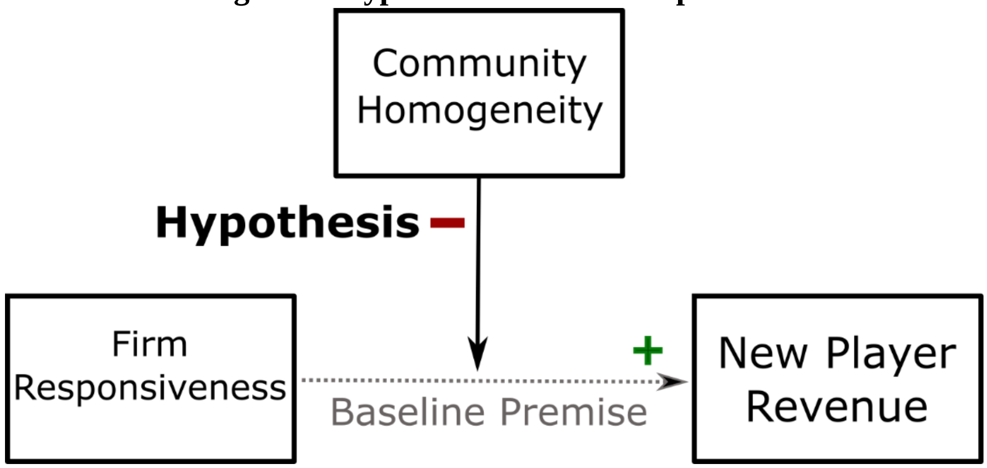


Such homogenous communities are more likely to embody niche preferences that are unrepresentative of the range of user preferences that characterizes the broader market (Shermon & Moeen, 2020). Insight gleaned from a homogenous community is therefore less likely to apply accurately to the broader market. As a result, experimentation in collaboration with a homogenous community may generate idiosyncratic and misleading demand signals that stunt future growth. Thus we hypothesize:

> *Hypothesis: Community homogeneity negatively moderates the relationship between firms' responsiveness to feedback and new customer revenue.*

Figure 2 summarizes the hypothesized relationships. The direct effect (the baseline premise) is represented by a dotted gray line; the hypothesis (the moderating effect) is represented by a solid dark line.

Figure 2: Hypothesis and baseline premise



3. Methods

3.1 Setting

We test our hypotheses using data on personal-computer (PC) video-game development. The video-game industry, which began in the 1970s as a niche market, has become an important economic sector. It is also the subject of a wide array of studies (Katila et al., 2022; Mollick, 2012; Ozcan & Eisenhardt, 2009; Rietveld & Eggers, 2018; Zhao et al., 2018; Zhu & Iansiti, 2012). In 2018, the industry reported projected global revenue of approximately $138 billion, an increase of 13% over the previous year (Wijman, 2018). Of this total, PC-game sales accounted for approximately 24%.

Video-game development consists of designing and coding electronic games. This process includes gameplay design, software engineering, and art design. Companies that develop video games range in size from small teams to large public firms like Electronic Arts. The broader ecosystem includes firms that publish games, develop hardware and software platforms, and provide development tools (Katila et al., 2022; Ozcan & Eisenhardt, 2009).

The video-game industry is an appropriate setting in which to study experimentation in user communities and innovation because it is characterized by a high level of market uncertainty (Katila et al., 2022). As one prominent developer aptly commented, *"Nobody knows how to always make a hit game. There is so much uncertainty."* This uncertainty puts a premium on experimentation—that is, on testing during development whether users will enjoy a given game and therefore purchase it. Furthermore, during development, design decisions pertaining to one element of a game often inadvertently affect other elements (Sylvester, 2013). This interdependence makes it difficult for developers to fully anticipate the consequences of design decisions: changing how high a character can jump can affect the type of terrain that characters can traverse, leading to problems with gameplay if such changes are unaccounted for in other parts of the game's design. As one developer explained, *"You can't [anticipate the consequences of design decisions]. How can you know exactly how tanks are going to affect the game until you put them in? We don't."*

To conduct our analysis, we collected data on video games hosted by Steam Early Access. Steam, the leading online digital distribution platform for PC games, was launched in 2003 by Valve, a successful video-game developer and publisher (Plunkett, 2013). Firms can list games on the Steam Store in exchange for up to a 30% share of sales revenue. In 2018, the Steam Store listed over 27,000 games for sale (Meer, 2019). Analysts estimate that PC game sales via Steam represent around 60% of global industry sales (Edwards, 2013; Hruska, 2018).

Early Access (EA), introduced by Steam in 2013, is a specialized channel that enables developers to experiment in conjunction with user communities while developing a PC game. Specifically, EA allows a game's developers to launch a game prototype on Steam; users can then buy and play the game, interact with one another, and provide feedback to the developer before the game's underlying code has been finalized (Welch, 2013). As is typical with the user communities that spring up around new products, many players choose to play such prototype games not because they are the most polished games available—they can buy fully developed games for that—but rather because they enjoy participating in development of a game. Developers in turn can benefit greatly from users who are willing to participate in the process of developing unfinished games. As one developer recounted: *"We launched [our game] in Early Access with about an hour of gameplay. It was enough to give the player a decent experience for their money right away. But it also gave us an opportunity to really incorporate [community] feedback as we continued to work on the game."*

The Steam Store lists EA games in a special category, where potential purchasers can read about why the developer chose to launch the game in EA and its current status. The developer is not formally obliged to maintain, change, or even complete the game; Valve thus encourages potential users to research each game and its developer thoroughly before purchase. A typical description reads:

> *The game is still undergoing very active development, so you can expect bugs and performance issues as the game progresses and changes. In its current state, it is more like a sandbox of technology and martial arts, and less like a story-driven action-adventure. If you want to play through a polished campaign, you may want to wait until the game is fully completed! If you would like to have fun experimenting with our technology and mechanics as they evolve, and support us as we complete the game, then now is a great time to join our community.*

Steam EA is an attractive setting for our research for two reasons. First, the purpose of EA is experimentation. Launching an incomplete game makes it possible for developers to test new ideas for a game's design. For example, a lead developer of a successful EA game explained that using the Steam community to experiment with new gameplay features prevented him from implementing what would have been a poor design decision: "*We added [a game feature], and it* *became clear that it was too hard and too boring for players. So we pivoted the following week based on that feedback and got rid of it."* Another developer reported using EA to *"fail fast and learn from the experience."*

Every EA game has its own user community—a collection of individuals with a shared interest in a particular EA game (Lakhani, 2016; Shah & Tripsas, 2007). Consistent with the literature on user communities (Puranam et al., 2014; Shah, 2005, 2006), anyone can join and participate in Steam EA communities (whether they play the game or not); people usually do so for intrinsic and non-financial reasons: they enjoy participating in the development process. Our interviews with users in EA communities (described in more depth below) pinpointed several motivations. As in other communities (Shah, 2006; Von Krogh et al., 2012), a common motivation is fun. In one user's words, *"Early Access games are a lot more fun to play. They are typically much more innovative. [Big budget, heavily marketed] titles are so incremental; they just improve the graphics or add a few new features to the same old formula."* Another motivation is social interaction (David & Shapiro, 2008). As one user explained: *"The community is what makes [Early Access games] fun to play for me. It's a social experience."* A third motivation is altruism (Osterloh & Rota, 2007; Von Krogh et al., 2012), which one user articulated: *"There is a certain joy in responding in forums and giving suggestions and feedback to developers, and then seeing your ideas being implemented."*

Also, as is typical in user communities (Shah & Tripsas, 2007), members of EA game user communities build on each other's ideas, typically via game-specific "community hubs" hosted on the Steam platform. Each community hub includes a space to upload content (e.g., screenshots and gameplay videos), a message board where users can discuss the game, and a space for users' game reviews. Users build on each other's insights via community discussions—whether about technical problems encountered during play, ideas for improving the game, or how to win. One developer commented: *"Sometimes if we don't explain things to someone, other players will."* Developers too participate in discussions, and rely on the community for ideas and feedback. As one observed: *"Listening to the community is extremely important. There were a couple of times where we ignored what the community was telling us, and I think that was a big mistake."*

3.2 Data

We constructed a panel dataset of Steam Early Access (EA) games launched between April 2015 (when data on products' commercial performance became available) and March 2018.[33] We excluded free games from analysis because the commercial-performance measure (revenue from new players) does not apply to free games. Also, consistent with other research (Mollick, 2012; Zhao et al., 2018), we excluded game expansions, or add-ons to existing games.

For the games remaining in the dataset, we aggregated time-varying data by calendar month. We chose months because interviews with game developers suggested that experimentation happens within this time frame. As one developer explained, *"We settled into a rhythm where we launch new features every month or so, and that gives us room to iterate. We can actually listen to our community and tweak things so that everything is working."*

Because our research design uses measures constructed using month-lagged and double-lagged game-review text, we excluded lagged and double-lagged game-months with fewer than 5 reviews. For each game, we observe each month during the time period spent in EA. For the panel analysis, we keep only firms with at least 2 observations. The result is an unbalanced panel of 370

[33] During this period, approximately 20% of Steam games were launched through EA (3,247 EA games vs. 12,804 standard games). Collectively, EA games closely resemble other games on such dimensions as genre and language. However, the majority (77%) of EA games were created by first-time developers (vs. 58% of regular games), and a larger portion were multiplayer (27% vs. 10% of regular games).

games over 2,754 game-months.

We collected quantitative data on Steam games from two primary sources: SteamDB (steamdb.info) and SteamSpy (steamspy.com). SteamDB is an online database that uses Steam's application-programming interface (API) to collect daily data on game characteristics, such as price, and on such developer activity as announcements and software updates. SteamSpy is another online database that uses Steam's API to estimate community characteristics and performance data for each game, such as the number of active users, average playtime, and total sales. SteamSpy's data has been used in several research studies (Bailey & Miyata, 2019; Brunt et al., 2020; Lin et al., 2019), and its accuracy has been confirmed by multiple game developers (Hall, 2015; Orland, 2015). See Appendix A for a detailed description of SteamSpy's data-collection process.

Finally, we supplemented these quantitative data with 29 first-hand interviews with PC game developers, industry executives, and community members; we also relied on archival interviews and blogs from Steam EA developers. The primary thrust of our interviews with developers and executives was twofold: (1) to understand the broader video-game industry, with a focus on the development process, experimentation, and the locus of uncertainty in product performance; and (2) to understand the specifics of PC game development on Steam, and especially when launching in EA. The interviews with EA game-community members focused on (1) why they participated in gaming communities and (2) why they purchased and played EA games. Overall, these interviews helped to validate empirical measures and improve interpretation of the results.

3.3 Measures

Dependent Variable. We operationalize **market growth** for each game i as revenue from new players in each month t (*New Player Revenue*$_{it}$). *New Player Revenue*$_{it}$ is measured as the increase in the number of game owners multiplied by the purchase price. Games are a one-time purchase, so customers can download and use the software freely for an unlimited period of time.[34] This is a primary indicator that game developers use to assess their own performance. We determined monthly units sold by calculating the average of SteamSpy's daily estimates of each game's number of owners over the course of each calendar month. We use the natural log of revenue to account for that measure's skewed distribution and normalize error residuals in our models.

Independent Variables. We operationalized **firm responsiveness to community feedback** ($Firm\ Responsiveness_{it-1}$) for each game *i* as the similarity between the text of the game's experimentation-related update announcements in month *t-1* and the text of user reviews in month *t-2*. To measure textual similarity, we use cosine similarity, a method for identifying textual similarity that has been used to compare similarities between papers and patents (Arts et al., 2018; Marx & Hsu, 2022), between R&D-project applications (Criscuolo et al., 2017), and between firms (Sharkey & Bromley, 2015). We argue that textual similarity between game updates and user reviews is an indicator that a firm is responding to comments submitted by the community of early users.

To create this measure, we collected the full textual corpus of developers' game-update announcements from SteamDB. For each game, in each month, we included only sentences that contained words associated with new features or changes to the game.[35] We then compared that text to the text of all game reviews in the prior month. To compute similarity between the two texts,

[34] In industry jargon, our measure captures only "base game sales," not including any revenue from subscription fees or extra-content sales (e.g., microtransactions).
[35] Specifically, we included only sentences containing any words that matched the following regular expression: "\\badd|\\bnew|\\bfeature|\\bchang|\\bnow\\b|\\bintro|no longer|experiment|\\btry|\\bis here\\b|rebalance." This list was compiled by reading more than 300 announcements from a diverse sample of developers to identify patterns associated with new game features, changes, introductions, and additions.

we removed stop-words (such as "and", "the", etc.), punctuation, and capitalization, stemmed the words, and then computed the TF-IDF weighting for each word, treating each review and announcement as its own document. We then computed the similarity between update announcements and reviews by computing the cosine similarity of the vector between TF-IDF weights for each text (Singhal, 2001).

The textual similarity between update announcements and previous reviews is an appropriate measure of firms' responses to community feedback because it captures when firms are directly addressing topics previously discussed by the community. The measure is in line with our conception of experimentation as a dynamic, ongoing process in which firms try out and test suggestions from an engaged community's feedback. Developers formulate a plan to improve the game based on feedback, then implement the feature or change gameplay accordingly.

In qualitative interviews, many developers confirmed that they had launched in Early Access to rapidly iterate, experiment (e.g., tweak game balance), and gauge the community's reactions to specific changes. In one developer's words: *"[In Early Access, we expect to] continuously develop new content on an ongoing basis, improving features, encouraging a healthy community feedback loop . . . keep analyzing the game and player behaviors, test ideas, and enhance or improve the service with minimal viable product iterations, boost winning ideas, and rethink or remove the ones that don't work as well."* We use a one-month lag for experimentation (and for other explanatory variables described below) both because it coincides with the monthly testing cadence described by developers and to partially address concerns about reverse causality.

Moderator Variable. We measured **community homogeneity** as the Herfindahl index of the distribution of market segments represented by all users who reviewed game *i* in month *t-2* ($Community\ homogeneity_{it-2}$). We used month $t-2$ because reviews from $t-2$ were used to calculate the firm's response in *t-1*. Formally, the degree of community homogeneity in each game is calculated as:

$$\sum_{i=1}^{n} s_i^2$$

where s_i represents the share of reviewers who belonged to market segment *i* out of *n* distinct segments.

Though any given user could be assigned to various market segments, we followed a common practice in market segmentation to assign each user to a single segment that most parsimoniously described them according to a set of both demographic and behavioral variables (e.g., "casual middle-age puzzle gamers" or "young high-intensity FPS gamers"). To find the market segment that best captured each user, we categorized all 4 million users from SteamDB into distinct market segments using a K-means clustering algorithm. The K-means algorithm is common in customer segmentation (Morwitz & Schmittlein, 1992; Yau & Tang, 2018). It iteratively attempts to find distinct non-overlapping subgroups for each datapoint in a way that minimizes variation within each group. We had access to 42 user-level variables: 31 game-category variables (e.g., strategy, first-person shooter) that captured the number of games in each category that the user owned, in addition to the average positivity of the user's reviews, the number of such reviews, mean words per review, percentage of games played that were in Early Access, proportion of games purchased directly from Steam, amount of time spent on the platform, number of games, % of games acquired for free, number of comments, number of "helpful" votes, and number of "funny" votes. Before applying the K-means algorithm, we used the UMAP algorithm to reduce the dimensionality of all the user-level variables from 42 down to 10 dimensions (McInnes et al., 2018). Applying the K-means algorithm to the dimensionally reduced data, we identified 5 distinct market segments—groups whose members were similar to each other and dissimilar to members of other groups. Having assigned each user to a market segment, we calculated homogeneity as the Herfindahl index of users' market segments, as described above.

Control Variables. We included several control variables. First, we controlled for the average **price** in each game-month ($Price_{it}$), since price can affect a customer's willingness to buy. We also controlled for **new game content,** using increases in the size (in megabytes) of each game *i* in month $t-1$ ($New\ content_{it-1}$). It was important that our measures for independent and moderating variables not be driven solely by how much people like a game. We controlled for **positivity** using the number of positive reviews divided by total reviews of game *i* in month *t-1* ($Positivity_{it-1}$).

Finally, because competition can suppress the commercial performance of the focal game, we controlled for **competition** ($Competition_{it}$). We measured competition using the number of rival games listed on Steam EA—that is, the number of games in the same genre as game *i* in month *t* (e.g., "action" and/or "simulation"). In the analysis, we logged these variables to adjust for skew and normalize the residuals of the model.

3.4 Statistical estimation

In line with our theory, we expect by changing the game in response to feedback from homogeneous communities, firms limit the game's appeal to new players and will have difficulty expanding. To estimate this effect, we applied linear fixed effect estimators to our monthly panel data. We also included game fixed effects to control for unobservable time-invariant heterogeneity between games, and month-year fixed effects to control for unobservable macro-environmental changes that affect game performance.

4. Results

4.1 Main Results

Table 1 reports descriptive statistics and Table 2 reports correlations. Table 3 presents the main results from our fixed-effect models.

Table 1: Descriptive statistics

Variable	Description	N	Mean	Std. dev. (overall)	Std. dev. (within)	Min	Med	Max
New Player Revenue$_{it}$	*Revenue from new players in month t (in US$ 1000s)*	2,754	338.90	4,974.35	2,533.64	0	30.25	157,348
Firm responsiveness$_{it-1}$	*Cosine similarity of text in update announcements in month t-1 to text in reviews in t-2*	2,754	0.16	0.11	11.48	0	0.15	1
Community homogeneity$_{it-2}$	*Herfindahl index of reviewers' market segments in month t-2*	2,754	0.33	0.09	0.07	0	0.31	1
Reviewers$_{it-1}$	*# of reviewers in month t-1*	2,754	114.84	623.15	216.80	5	29	15,755
Players$_{it-1}$	*# of players in month t-1 (in 1000s)*	2,754	27.85	345.79	221.29	0	3.32	13,413
Positivity$_{it-1}$	*# of "up" reviews / # of total reviews in month t-1*	2,754	0.80	0.17	0.10	0	0.83	1
New content$_{it-1}$	*Increase in game size in month t-1 (in MBs)*	2,754	214.36	1,227.17	1,232.72	0	5	55,254
Price$_{it}$	*Mean price of game (in US$) in month t*	2,754	17.10	8.19	1.86	0	15.80	54
Competition$_{it}$	*# of games in same category in month t*	2,754	240.81	377.76	76.52	0	90	1,919

Table 2: Correlation matrix

	1	2	3	4	5	6	7	8	9
1 New Player Revenue$_{it}$	1								
2 Firm responsiveness$_{it-1}$	0.017	1							
3 Community homogeneity$_{it-2}$	-0.016	-0.139	1						
4 Reviewers$_{it-1}$	0.913	0.077	-0.058	1					
5 Players$_{it-1}$	0.954	0.017	-0.025	0.872	1				
6 Positivity$_{it-1}$	-0.051	0.061	0.070	-0.018	-0.049	1			
7 New content$_{it-1}$	-0.004	0.074	-0.031	-0.0001	-0.005	-0.032	1		
8 Price$_{it}$	0.082	0.205	-0.087	0.116	0.074	-0.129	0.151	1	
9 Competition$_{it}$	-0.028	-0.074	0.001	-0.029	-0.025	0.041	0.009	-0.079	1

Table 3. Main results

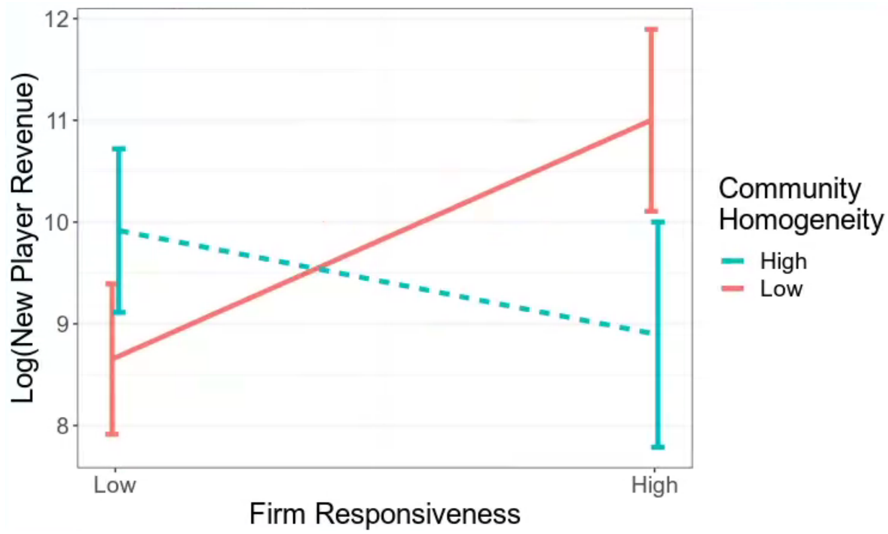
	Dependent variable: *New Player Revenue$_{it}$*		
	(1)	**(2)**	**(3)**
Firm responsiveness$_{it-1}$		0.20	0.17
		(0.07)	(0.08)
Community homogeneity$_{it-2}$		-0.06	-0.11
		(0.09)	(0.09)
Firm responsiveness$_{it-1}$ ×			-0.21
Community homogeneity$_{it-2}$			(0.09)
Reviewers$_{it-1}$	0.52	0.46	0.48
	(0.17)	(0.17)	(0.17)
Players$_{it-1}$	0.57	0.58	0.57
	(0.18)	(0.18)	(0.18)
Positivity$_{it-1}$	0.95	0.93	0.95
	(0.70)	(0.70)	(0.70)
New content$_{it-1}$	-0.00	-0.01	-0.01
	(0.02)	(0.02)	(0.02)
Price$_{it}$	0.58	0.64	0.61
	(0.84)	(0.81)	(0.82)
Competition$_{it}$	0.02	0.02	0.04
	(0.41)	(0.41)	(0.42)
Game FE	YES	YES	YES
Month-year FE	YES	YES	YES
Adj. R^2	0.35	0.35	0.36
Within R^2	0.04	0.04	0.04
N	2754	2754	2754

Notes. Robust standard errors clustered at the game-level appear in parentheses. To ease interpretation, the independent and moderator variables are standardized by mean-centering and z-scoring. All control variables were log-transformed to normalize the model residuals (except *Positivity$_{it-1}$*, which was already normal). All models include game-level and month-year-level fixed effects.

Table 3 tests our hypothesis with three models that use (logged) *New Player Revenue*$_{it}$ as the dependent variable. Model 1 includes only control variables, along with game and month-year fixed effects. As expected, *Reviewers*$_{it-1}$ and *Players*$_{it-1}$, are positively related to *New Player Revenue*$_{it}$. *Positivity*$_{it-1}$, *Price*$_{it-1}$ and *Competition*$_{it-1}$ are null. Model 2, which adds the independent and moderator variables, indicates that a standard deviation increase in *Firm responsiveness*$_{it-1}$ is associated with an approximately 20% increase in *New Player Revenue*$_{it}$. This result is consistent with the baseline premise of our model, and is consistent with prior work. As expected, the coefficient on *Community homogeneity*$_{it-2}$ alone indicates a null effect. The lack of a significant result is reassuring: it suggests that we have effectively disentangled the community's degree of homogeneity from its size (which would be strongly positively correlated with *New Player Revenue*$_{it}$).

To formally test our hypothesis, Model 3 adds the interaction effect *Firm Responsiveness*$_{it-1}$ $\times$ *Community homogeneity*$_{it-2}$. The significant coefficient estimate indicates that, for games with a value of *Community homogeneity*$_{it-2}$ a standard deviation above the mean, a standard deviation increase in *Firm responsiveness*$_{it-1}$ is associated with about a 20% decrease in *New Player Revenue*$_{it}$. For ease of interpretation, the predicted marginal effects of this result are displayed in Figure 3. This pattern confirms our hypothesis that community homogeneity negatively moderates the relationship between firm responsiveness and products' commercial performance.

Figure 3. Marginal effects from main model



Notes. This figure presents predicted marginal effects from the main result (Model 3 in Table 3). "High" and "Low" represent 2 standard deviations above and below the mean, respectively.

4.2 Robustness check: Inverse Probability Treatment Weighting and Winsorization

Though fixed effects effectively control for time-invariant inter-game heterogeneity, endogeneity concerns may persist if our treatment (*Firm responsiveness*$_{it-1}$) is not randomly assigned within firms over time. We address this issue using inverse probability treatment weighting (IPTW), which generalizes propensity score matching to panel data (Azoulay et al., 2009; Wang et al., 2021). Certain features of developers or of game projects may predispose them simultaneously to be more responsive to communities and to experience market growth. This endogeneity issue is mitigated by weighting each observation by the inverse of its probability of treatment, as a function of potentially confounding observables. For a full explanation of how we applied this procedure, including covariate balances pre- and post-weighting, see Appendix B.

The results of the IPTW analysis are consistent with our main results. Table 4, model 1, repeats the estimation from Table 3 but adds the IPTW weights to the estimation. The results are even more dramatic. The positive effect of *Firm responsiveness*$_{it-1}$ is smaller, and the negative effect of the interaction term *Firm responsiveness*$_{it-1}$ × *Community homogeneity*$_{it-2}$ is larger than in the original model. Therefore, when accounting for observable selection into *Firm responsiveness*$_{it-1}$, firms are more likely to see responsiveness negatively impact commercial performance than in the original model.

Finally, we check that extreme values are not driving our results. Table 4, model 2, repeats the same estimation as model 1, but Winsorizes all right-hand-side variables at the 5% level. This model indicates an even larger negative estimate of the interaction term *Firm responsiveness*$_{it-1}$ × *Community homogeneity*$_{it-2}$. In sum, after accounting for observable selection effects and extreme values, the results remain consistent and even stronger than in the baseline models.

Table 4. Robustness checks: IPTW and Winsorization

Dependent variable: ***New Player Revenue***$_{it}$	**(1)**	**(2)**
Firm responsiveness$_{it-1}$	0.15	0.14
	(0.09)	(0.11)
Community homogeneity$_{it-2}$	-0.07	-0.09
	(0.10)	(0.12)
Firm responsiveness$_{it-1}$ ×	-0.25	-0.35
Community homogeneity$_{it-2}$	(0.12)	(0.14)
Reviewers$_{it-1}$	0.63	0.66
	(0.18)	(0.20)
Players$_{it-1}$	0.48	0.50
	(0.20)	(0.22)
Positivity$_{it-1}$	1.94	2.33
	(0.73)	(0.84)
New content$_{it-1}$	0.01	0.00
	(0.02)	(0.02)
Price$_{it}$	1.08	0.78
	(0.69)	(1.72)
Competition$_{it}$	0.27	0.25
	(0.63)	(0.64)
IPTW weights	YES	YES
5% Winsorization	NO	YES
Game FE	YES	YES
Month-Year FE	YES	YES
Adj. R^2	0.40	0.39
Within R^2	0.06	0.05
N	2614	2614

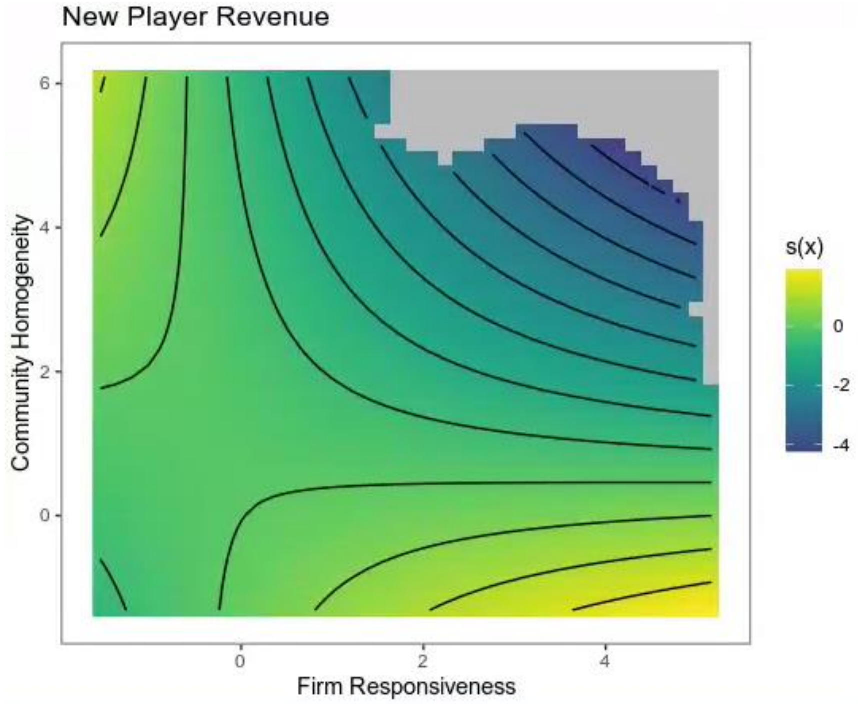
Notes. Robust standard errors clustered at the game-level appear in parentheses. To ease interpretation, the independent and moderator variables are standardized by mean-centering and z-scoring. All control variables were log-transformed to normalize the model residuals (except *Positivity*$_{it-1}$). All models include game-level and month-year-level fixed effects. Model 1 uses IPTW weights; Model 2 uses IPTW weights and 5% Winsorization on all the right-hand-side variables.

4.3 Nonlinear visualizations of interaction terms

After testing our hypotheses using the linear models presented above, we also used Generalized Additive Models (GAMs) as a post hoc analysis to validate that the interaction effects were not merely the result of linear modeling assumptions (Hastie et al., 2009). GAMs allow us to validate the interaction term effects with minimal functional form assumptions, balancing the bias-variance tradeoff to avoid overfitting the data (Choudhury et al., 2021).

Figure 4 displays the estimated interaction effects. The GAM model included all the variables displayed in Table 3, but model the interaction term using a tensor smooth term that allows for more complex nonlinear interaction effects. The tensor smooth term was strongly statistically significant. The GAM model in Figure 4 demonstrates that, controlling for other factors, when *Community homogeneity*$_{it-2}$ was low, increases in *Firm Responsiveness*$_{it-1}$ were associated with more *New Player Revenue*$_{it}$. However, the opposite was true when *Community homogeneity*$_{it-2}$ was high. As with the linear models, these visualizations validate the same findings as the linear models presented above.

Figure 4. 2D Smooth plots for generalized additive models (GAMs)



Notes. This figure presents two-dimensional smooth plots from generalized additive models (GAMs), each of which includes all the independent variables and control variables listed in Table 3. The dependent variable is *New Player Revenue*$_{it}$; the model uses a tensor product smooth to capture interaction effects between *Firm responsiveness*$_{it-1}$ and *Community homogeneity*$_{it-2}$. Higher predicted revenue is represented by yellow, lower is represented by dark blue.

4.4 Additional analysis: novel product vs. non-novel product subsamples

We have shown that responding to feedback from homogenous communities can limit market growth. How do we reconcile this with prior research demonstrating that large companies often have difficulty launching breakthrough products precisely because they are too focused on the mainstream market (Allen, 2022b, 2022a; Christensen & Bower, 1996; Kapoor & Klueter, 2015; Vinokurova & Kapoor, 2020)?

We suggest that targeting a niche of early adopters could be a more successful strategy for relatively novel products, for which a completely new value proposition limits initial appeal to a small fraction of people, but where there is the possibility of going viral. By contrast, our theory is more applicable to products that are attempting to penetrate the market with a known value proposition. For these products, catering to niche preferences may be a less effective strategy. This proposition would align with findings that rational updating based on external information is more effective under conditions of lower uncertainty (Kapoor & Wilde, 2022).

We conducted additional analyses that confirm that our result is driven by products that are relatively familiar to the market, rather than novel. We operationalize novel games as those that are categorized differently than previously existing games in Steam. We are able to see how games are categorized using Steam Tags. There are 351 total tags, which users use to categorize games (e.g., Turn-based tactics, Heist, Minigames, Dystopian, Puzzle, Magic, etc.). If games attract enough user votes for a tag, it is officially used as a tag for the game. About half of games have 5 or fewer tags, but some have up to 21. We defined games as "novel" if their top 3 tags were a unique combination relative to existing games. We used the top 3 tags because these represent the most important categorizations according to users, and because using all tags would classify most games as novel (not useful). Among the full set of about 24,000 Steam games with tags, our method categorized 23% as novel.

Table 5 repeats the main analysis using our subset of Early Access games, divided into novel and non-novel subsamples. Column 1 reports the estimation on the novel subsample. The interaction term estimates a null result. Column 2 reports the estimation on the non-novel subsample. The interaction term is significantly negative, as in the main analysis. This suggests that our result is indeed driven by non-novel products.

Table 5. Subsample analysis: Novel vs. non-novel products

Dependent variable: ***New Player Revenue**$_{it}$*	**(1) Novel Product Subsample**	**(2) Non-novel Product Subsample**
Firm responsiveness$_{it-1}$	0.17	0.17
	(0.12)	(0.10)
Community homogeneity$_{it-2}$	-0.20	-0.05
	(0.15)	(0.11)
Firm responsiveness$_{it-1}$ ×	-0.12	-0.23
Community homogeneity$_{it-2}$	(0.16)	(0.10)
Reviewers$_{it-1}$	0.50	0.48
	(0.31)	(0.19)
Players$_{it-1}$	0.66	0.43
	(0.33)	(0.20)
Positivity$_{it-1}$	0.85	1.10
	(0.95)	(0.93)
New content$_{it-1}$	0.02	-0.03
	(0.04)	(0.03)
Price$_{it}$	-2.35	1.17
	(1.83)	(0.59)
Competition$_{it}$	0.55	-0.21
	(0.52)	(0.59)
Game FE	YES	YES
Month-Year FE	YES	YES
Adj. R^2	0.36	0.35
Within R^2	0.05	0.04
N	1122	1632

Notes. Robust standard errors clustered at the game-level appear in parentheses. To ease interpretation, the independent and moderator variables are standardized by mean-centering and z-scoring. All control variables were log-transformed to normalize the model residuals (except *Positivity$_{it-1}$*). All models include game-level and month-year-level fixed effects. Model 1 uses a subsample of novel products; Model 2 uses a subsample of non-novel products.

5. Discussion

This paper scrutinizes the circumstances in which experimentation with user communities enhances (or diminishes) an innovation's commercial growth. By means of a large-sample archival study of PC game development, supplemented with interviews, we empirically analyzed how community homogeneity moderates the relationship between firms' responsiveness to community feedback and their revenue from new users. Our results suggest an overlooked contingency that provides new insight for innovation theorists.

Although it is common for communities to be characterized by a homogenous set of users, we find that community homogeneity strongly negatively moderates the relationship between firms' responsiveness and their products' growth in the market. That is, firms' responsiveness to feedback from a homogenous community decreased revenue from new users; responsiveness to more-representative communities increased revenue from new users. This section will address the theoretical and practical implications of these results and discuss our study's contributions to research on user communities, firm innovation, and business experimentation.

5.1 Evidence that experimentation in user communities (usually) aids commercial performance

It is a key premise of this study, drawn from our review of prior work, that user communities help firms innovate via experimentation. Indeed, we found that firms can improve their commercial performance via experimentation in cooperation with user communities. Specifically, communities seem to provide insight into customer preferences. Firms glean such insight by observing users using their products. A Steam EA developer explained: *"The times I learn the most about my games are when I get them in the hands of real players and watch what they do and* *how they react."* User communities also offer a forum for social interactions in which users build on each other's insights and ultimately provide better information to firms. Another developer described gaining insight from reading community message boards: *"Sometimes you just hear the same old gripes when you read through [a community message board]. But, in fact, a lot of the features in the latest build were inspired by the discussions on there."* In short, information from user communities guides improvements in products' design and technical performance that result in greater commercial success. But there are limits.

5.2 A constraint on the benefits of user-community experimentation: community homogeneity

Existing research on firms' community-driven experimentation has emphasized drawing on a community highly engaged in feedback (Harhoff & Lakhani, 2016). We have taken issue with this insufficiently nuanced perspective by asking under what conditions experimentation with user communities enhances innovation. We have identified an important contingency specific to experimenting in new-product development in conjunction with user communities. In short, we have shown that homogenous user communities can mislead experimentation-driven new-product development.

How can homogenous communities misdirect innovation? Our findings suggest that a more-homogenous community tends to draw on more uniform and less broadly represented market preferences. Catering to these niche preferences can increasingly restrict the commercial appeal of the product. The lead developer of a critically acclaimed Steam EA game explained how critical it is to observe many users when experimenting in EA:

> *If you watch enough different types of players—you watch the skilled players, the beginners, the ones who are really creative— . . . you'll learn by sampling a wide range of data. What are the things that come up over and over? What are the things that just came up once and*

you don't actually have to worry about? And that saves you from focusing on something that is coming from just one person.

This paper specifies boundary conditions within which firms are likely to realize benefits from experimentation with user communities. Much prior work extols the virtues of engaged user communities (Antorini et al., 2012; Baldwin & Von Hippel, 2011; Bayus, 2013; Hienerth et al., 2014). Our study joins a growing body of research that takes exception to across-the-board assertions of user communities' beneficial role in innovation (Bremner & Eisenhardt, 2021; Shah & Nagle, 2020). But unlike prior research on how *firms'* and *users'* characteristics (Anthony et al., 2009) shape the insights that firms gain from user feedback (Foss et al., 2011), this paper provides a complementary view of how a *community* characteristic—homogeneity—moderates the experimentation–innovation relationship.

5.3 A community engagement dilemma?

Our results highlight a previously overlooked tension of engaging with user communities for innovation: an overly facile interpretation of our finding might prompt firms to ignore feedback from small, engaged user communities, because acting on their feedback to modify an inchoate product risks undermining its commercial potential. But such communities are valued, in part, for their voluntary work (Shah & Nagle, 2020), freely sharing of ideas (Shah & Nagle, 2020; Shah & Tripsas, 2007), and self-selection into tasks (Lakhani, 2016; O'Mahony, 2003; Shah, 2006)—all of which arise from members' strong intrinsic motivation to contribute (Boudreau & Lakhani, 2015; Lakhani, 2016). Ignoring engagement from ardent users risks discouraging their motivation to engage in the future, all but ensuring that their communities never become large, valuable resources in commercialization (Bapna et al., 2019; A. Murray et al., 2020; A. Murray & Fisher, 2022; Piezunka & Dahlander, 2015, 2019). This presents a dilemma that suggests a need for heightened sensitivity to the dual nature of collective engagement (and to the underlying motivations that drive it), and perhaps for more careful screening of well-intended ideas and solutions that emanate from small communities. More research is needed to better understand the extent of this dilemma and ways to resolve it.

5.4 Managerial implications

Generally speaking, firms can increase the market for their new products by experimenting in conjunction with user communities. But the nature of the community matters. Our findings suggest a need for heightened sensitivity to the well-intentioned ideas and solutions that emanate generated by homogenous communities.

Experimentation is beneficial to firms when the participating communities are representative of the market outside the community, but not necessarily when they are highly homogenous. The implication of this finding for managers is that user communities can offer valuable insight to product-design and engineering efforts, but that catering to a niche of highly engaged users entails risks. As one developer warned: *"Listen to your community. But I would also say: know when not to listen."*

Yet at the same time, incorporating feedback from the community is central to achieving a large and valuable community resource. Firms that experiment in homogenous communities face a dilemma: incorporate the feedback and risk building a product with narrow appeal, or ignore the feedback and lose the support of the community.

5.5 Limitations and scope conditions

As is true of all research, our findings have potential alternative explanations. One is that the amount of customer feedback, firms' responsiveness, and commercial performance could all be correlated with games that intentionally appeal to broader audiences. But this explanation does not account for the fact that our analysis relies primarily on within-firm variation. Another alternative explanation is that firms' responsiveness enhances commercial performance and engagement regardless of actual changes to the game, simply because users like a highly responsive firm. Though plausible, this reasoning is inconsistent with our analysis, which does not demonstrate a universal link between responsiveness and success; instead it demonstrates a tension between engagement with homogenous communities and near-term performance.

Our findings point to several avenues for future research. First, our finding that user communities can derail an innovation's commercial performance highlights that user communities are just one group of stakeholders amongst other internal and external groups, each with different values and objectives. Future research could explore how internal organizational structure changes how firms respond to user communities' input and influence (Eklund & Kapoor, 2022; Kapoor & Klueter, 2015; Lee, 2022; Vinokurova & Kapoor, 2020), or explore how firms balance the demands of user communities with the demands of other external audiences that shape innovation incentives (Benner & Ranganathan, 2012, 2013; McDonald & Allen, 2021; McDonald & Gao, 2019). Second, our additional analysis hinted that our results are driven by non-novel innovations, suggesting that another fruitful avenue would be to examine how engaging user communities with different attributes leads to different outcomes in other contexts. Nascent industries may present a particularly interesting context (Gao & McDonald, 2022; Moeen et al., 2020), wherein niche communities of early adopters may be even more important for resolving uncertainty. Relatedly, it would be important to further explore the idea that highly engaged homogenous communities may degrade performance for "safe" or incremental innovations, but encourage breakthrough innovations that may take longer to diffuse (Allen, 2022a, 2022c; Christensen et al., 2018; Vinokurova & Kapoor, 2020). For now, we leave these questions to future research.

6. Conclusion

Prior research suggests that user communities are central to innovation, but leaves largely unexplored the impact of community-level attributes on firm-led innovation efforts. We address this gap using data on experimentation in nearly 400 PC-game-development projects. Our theory and empirical evidence confirm that experimentation jointly with user communities positively influences the resulting products' commercial performance. We also find, however, that the degree of a community's homogeneity powerfully moderates the benefits of experimentation: firms' innovation performance improves when they solicit insight from a representative community of users but suffers when engaging a homogenous community that is concentrated in a narrow market niche.

References

Afuah, A., & Tucci, C. L. (2012). Crowdsourcing as a solution to distant search. *Academy of Management Review*, *37*(3), 355–375.

Allen, R. T. (2022a). *Methodological Pluralism and Innovation in Data-driven Organizational Cultures*.

Allen, R. T. (2022b). The Market-Size Paradox of Niche Market Innovations. In *Academy of Management Proceedings* (Vol. 2022, Issue 1). Academy of Management Briarcliff Manor, NY 10510.

Allen, R. T. (2022c). The Market Size Paradox of Niche Product Innovations. *Academy of Management Proceedings*, *2022*(1), 17600.

Anthony, D., Smith, S. W., & Williamson, T. (2009). Reputation and reliability in collective goods: The case of the online encyclopedia Wikipedia. *Rationality and Society*, *21*(3), 283–306.

Antorini, Y. M., Muñiz Jr, A. M., & Askildsen, T. (2012). Collaborating with customer communities: Lessons from the LEGO Group. *MIT Sloan Management Review*, *53*(3), 73.

Arts, S., Cassiman, B., & Gomez, J. C. (2018). Text matching to measure patent similarity. *Strategic Management Journal*, *39*(1), 62–84.

Autio, E., Dahlander, L., & Frederiksen, L. (2013). Information exposure, opportunity evaluation, and entrepreneurial action: An investigation of an online user community. *Academy of Management Journal*, *56*(5), 1348–1371.

Azoulay, P., Ding, W., & Stuart, T. (2009). The impact of academic patenting on the rate, quality and direction of (public) research output. *Journal of Industrial Economics*, *57*(4), 637–676.

Bailey, E., & Miyata, K. (2019). Improving video game project scope decisions with data: An analysis of achievements and game completion rates. *Entertainment Computing*, *31*, 100299.

Baldwin, C., Hienerth, C., & von Hippel, E. A. (2006). How user innovations become commercial products: A theoretical investigation and case study. *Research Policy*.

Baldwin, C., & Von Hippel, E. (2011). Modeling a paradigm shift: From producer innovation to user and open collaborative innovation. *Organization Science*, *22*(6), 1399–1417.

Bapna, S., Benner, M. J., & Qiu, L. (2019). Nurturing Online Communities: An Empirical Investigation. *MIS Quarterly*, *43*(2).

Bayus, B. L. (2013). Crowdsourcing new product ideas over time: An analysis of the Dell IdeaStorm community. *Management Science*, *59*(1), 226–244.

Benner, M. J., & Ranganathan, R. (2012). Offsetting Illegitimacy? How Pressures From Securities Analysts Influence Incumbents in the Face of New Technologies. *Academy of Management Journal*, *55*(1), 213–233.

Benner, M. J., & Ranganathan, R. (2013). Divergent reactions to convergent strategies: Investor beliefs and analyst reactions during technological change. *Organization Science*, *24*(2), 378–394.

Bingham, C. B., & Davis, J. P. (2012). Learning sequences: Their existence, effect, and evolution. *Academy of Management Journal*, *55*(3), 611–641.

Boudreau, K. J., & Lakhani, K. R. (2015). “Open” disclosure of innovations, incentives and follow-on reuse: Theory on processes of cumulative innovation and a field experiment in computational biology. *Research Policy*, *44*(1), 4–19.

Bremner, R. P., & Eisenhardt, K. M. (2021). Organizing form, experimentation, and

performance: Innovation in the nascent civilian drone industry. *Organization Science*.
Bresnahan, T. F., Davis, J. P., & Yin, P.-L. (2014). Economic value creation in mobile applications. In *The changing frontier: Rethinking science and innovation policy* (pp. 233–286). University of Chicago Press.
Brown, S. L., & Eisenhardt, K. M. (1997). The art of continuous change: Linking complexity theory and time-paced evolution in relentlessly shifting organizations. *Administrative Science Quarterly*, 1–34.
Brunt, C. S., King, A. S., & King, J. T. (2020). The influence of user-generated content on video game demand. *Journal of Cultural Economics*, *44*(1), 35–56.
Camuffo, A., Cordova, A., Gambardella, A., & Spina, C. (2019). A Scientific Approach to Entrepreneurial Decision Making: Evidence from a Randomized Control Trial. *Management Science*, *October*.
Cao, R., Koning, R., & Nanda, R. (2021). *Biased Sampling of Early Users and the Direction of Startup Innovation*.
Chatterji, A. K., & Fabrizio, K. (2012). How do product users influence corporate invention? *Organization Science*, *23*(4), 971–987.
Chatterji, A. K., & Fabrizio, K. R. (2014). Using users: When does external knowledge enhance corporate product innovation? *Strategic Management Journal*, *35*(10), 1427–1445.
Choudhury, P., Allen, R. T., & Endres, M. G. (2021). Machine Learning for Pattern Discovery in Management Research. *Strategic Management Journal*.
Christensen, C. M., & Bower, J. L. (1996). Customer Power, Strategic Investment, and the Failure of Leading Firms. *Strategic Management Journal*, *17*(3), 197–218.
Christensen, C. M., McDonald, R., Altman, E. J., & Palmer, J. E. (2018). Disruptive Innovation: An Intellectual History and Directions for Future Research. *Journal of Management Studies*.
Criscuolo, P., Dahlander, L., Grohsjean, T., & Salter, A. (2017). Evaluating novelty: The role of panels in the selection of R&D Projects. *Academy of Management Journal*, *60*(2), 433–460.
Dahlander, L., & Frederiksen, L. (2012). The core and cosmopolitans: A relational view of innovation in user communities. *Organization Science*, *23*(4), 988–1007.
Dahlander, L., & Gann, D. M. (2010). How open is innovation? *Research Policy*, *39*(6), 699–709.
Dahlander, L., & Magnusson, M. G. (2005). Relationships between open source software companies and communities: Observations from Nordic firms. *Research Policy*, *34*(4), 481–493.
Dahlander, L., & Piezunka, H. (2014). Open to suggestions: How organizations elicit suggestions through proactive and reactive attention. *Research Policy*, *43*(5), 812–827.
Dahlander, L., & Wallin, M. W. (2006). A man on the inside: Unlocking communities as complementary assets. *Research Policy*, *35*(8), 1243–1259.
David, P. A., & Shapiro, J. S. (2008). Community-based production of open-source software: What do we know about the developers who participate? *Information Economics and Policy*, *20*(4), 364–398.
Edwards, C. (2013). Valve lines up console partners in challenge to Microsoft, Sony. *Bloomberg Business. Available Online at: Http://Www. Bloomberg. Com/News/Articles/2013-11-04/Valve-Lines-up-Consolepartners-in-Challenge-to-Microsoft-Sony*.
Eklund, J., & Kapoor, R. (2022). Mind the Gaps: How Organization Design Shapes the Sourcing of Inventions. *Organization Science*, *33*(4), 1319–1339.

Felin, T., Gambardella, A., Stern, S., & Zenger, T. (2019). Lean startup and the business model: Experimentation revisited. *Long Range Planning*, *March*, 101889.
Felin, T., & Zenger, T. R. (2014). Closed or open innovation? Problem solving and the governance choice. *Research Policy*, *43*(5), 914–925.
Foss, N. J., Laursen, K., & Pedersen, T. (2011). Linking customer interaction and innovation: The mediating role of new organizational practices. *Organization Science*, *22*(4), 980–999.
Franke, N., & Shah, S. (2003). How communities support innovative activities: an exploration of assistance and sharing among end-users. *Research Policy*, *32*(1), 157–178.
Franz, K. G. (1999). *Narrating automobility: Travelers, tinkerers, and technological authority in the twentieth century*. Brown University.
Gao, C., & McDonald, R. (2022). Shaping nascent industries: Innovation strategy and regulatory uncertainty in personal genomics. *Administrative Science Quarterly*.
Hall, C. (2015). *Steam Spy scrapes Steam user accounts to estimate sales data. Polygon*.
Harhoff, D., & Lakhani, K. R. (2016). *Revolutionizing innovation: Users, communities, and open innovation*. Mit Press.
Hastie, T., Tibshirani, R., Friedman, J. H., & Friedman, J. H. (2009). *The elements of statistical learning: data mining, inference, and prediction* (Vol. 2). Springer.
Hienerth, C., Von Hippel, E., & Jensen, M. B. (2014). User community vs. producer innovation development efficiency: A first empirical study. *Research Policy*, *43*(1), 190–201.
Hruska, J. (2018). Steam Earned an Estimated $4.3B in 2017, but Benefits Flow to Handful of Titles. *Extreme Tech*.
Jeppesen, L. B., & Molin, M. J. (2003). Consumers as co-developers: Learning and innovation outside the firm. *Technology Analysis & Strategic Management*, *15*(3), 363–383.
Kapoor, R., & Klueter, T. (2015). Decoding the adaptability-rigidity puzzle: Evidence from pharmaceutical incumbents' pursuit of gene therapy and monoclonal antibodies. *Academy of Management Journal*, *58*(4), 1180–1207.
Kapoor, R., & Wilde, D. (2022). Peering into a crystal ball: Forecasting behavior and industry foresight. *Strategic Management Journal*.
Katila, R., Piezunka, H., Reineke, P., & Eisenhardt, K. M. (2022). Big fish versus big pond? Entrepreneurs, established firms, and antecedents of tie formation. *Academy of Management Journal*, *65*(2), 427–452.
Katila, R., Thatchenkery, S., Christensen, M. Q., & Zenios, S. (2017). Is there a doctor in the house? Expert product users, organizational roles, and innovation. *Academy of Management Journal*, *60*(6), 2415–2437.
Lakhani, K. R. (2016). Managing communities and contests to innovate with crowds. *Revolutionizing Innovation: Users, Communities, and Open Innovation*, *109*.
Lakhani, K. R., & Von Hippel, E. (2003). How open source software works: "free" user-to-user assistance. *Research Policy*, *32*(6), 923–943.
Lee, S. (2022). The myth of the flat start-up: Reconsidering the organizational structure of start-ups. *Strategic Management Journal*, *43*(1), 58–92.
Lifshitz-Assaf, H. (2018). Dismantling knowledge boundaries at NASA: The critical role of professional identity in open innovation. *Administrative Science Quarterly*, *63*(4), 746–782.
Lilien, G. L., Morrison, P. D., Searls, K., Sonnack, M., & Hippel, E. von. (2002). Performance assessment of the lead user idea-generation process for new product development. *Management Science*, *48*(8), 1042–1059.
Lin, D., Bezemer, C.-P., Zou, Y., & Hassan, A. E. (2019). An empirical study of game reviews

on the Steam platform. *Empirical Software Engineering*, *24*(1), 170–207.
MacCormack, A., Verganti, R., & Iansiti, M. (2001). Developing products on "Internet time": The anatomy of a flexible development process. *Management Science*, *47*(1), 133–150.
Marx, M., & Hsu, D. H. (2022). Revisiting the entrepreneurial commercialization of academic science: Evidence from "Twin" discoveries. *Management Science*, *68*(2), 1330–1352.
McDonald, R. M., & Allen, R. T. (2021). A Spanner in the Works: Category-Spanning Entrants and Audience Valuation of Incumbents. *Strategy Science*, *May*.
McDonald, R. M., & Eisenhardt, K. M. (2020). Parallel play: Startups, nascent markets, and effective business-model design. *Administrative Science Quarterly*, *65*(2), 483–523.
McDonald, R. M., & Gao, C. (2019). Pivoting Isn't Enough? Managing Strategic Reorientation in New Ventures. *Organization Science*, *November*.
McInnes, L., Healy, J., & Melville, J. (2018). Umap: Uniform manifold approximation and projection for dimension reduction. *ArXiv Preprint ArXiv:1802.03426*.
Meer, A. (2019). Half of all games on steam came out since 2017. *Rock Paper Shotgun. Available at: Https://Www. Rockpapershotgun. Com/2019/01/15/How-Many-Games-Are-on-Steam/(Accessed 19 December 2019)*.
Miner, A. S., Bassof, P., & Moorman, C. (2001). Organizational improvisation and learning: A field study. *Administrative Science Quarterly*, *46*(2), 304–337.
Moeen, M., Agarwal, R., & Shah, S. K. (2020). Building industries by building knowledge: Uncertainty reduction over industry milestones. *Strategy Science*, *5*(3), 218–244.
Mollick, E. (2012). People and process, suits and innovators: The role of individuals in firm performance. *Strategic Management Journal*, *33*(9), 1001–1015.
Mollick, E. (2016). Filthy lucre? Innovative communities, identity, and commercialization. *Organization Science*, *27*(6), 1472–1487.
Moore, G. A. (1991). *Crossing the chasm*.
Morwitz, V. G., & Schmittlein, D. (1992). Using segmentation to improve sales forecasts based on purchase intent: Which "intenders" actually buy? *Journal of Marketing Research*, *29*(4), 391–405.
Murray, A., & Fisher, G. (2022). When More Is Less: Explaining the Curse of Too Much Capital for Early-Stage Ventures. *Organization Science*.
Murray, A., Kotha, S., & Fisher, G. (2020). Community-based resource mobilization: How entrepreneurs acquire resources from distributed non-professionals via crowdfunding. *Organization Science*, *31*(4), 960–989.
Murray, F., & Tripsas, M. (2004). The exploratory processes of entrepreneurial firms: The role of purposeful experimentation. In *Business strategy over the industry lifecycle*. Emerald Group Publishing Limited.
Nagle, F. (2018). Learning by Contributing : Gaining Competitive Advantage Through Contribution to Crowdsourced Public Goods Learning by Contributing : Gaining Competitive Advantage Through Contribution to Crowdsourced Public Goods. *Organization Science*, *June*.
Nagle, F. (2019). Open source software and firm productivity. *Management Science*, *65*(3), 1191–1215.
O'Mahony, S. (2003). Guarding the commons: how community managed software projects protect their work. *Research Policy*, *32*(7), 1179–1198.
Orland, K. (2015). *Introducing Steam Gauge: Ars reveals Steam's most popular games. ars technica*.

Osterloh, M., & Rota, S. (2007). Open source software development—Just another case of collective invention? *Research Policy, 36*(2), 157–171.

Ozcan, P., & Eisenhardt, K. (2009). Origin of alliance portfolios: Entrepreneurs, network strategies, and firm performance. *Academy of Management Journal*, *52*(2), 246–279.

Piezunka, H., & Dahlander, L. (2015). Distant search, narrow attention: How crowding alters organizations' filtering of suggestions in crowdsourcing. *Academy of Management Journal*, *58*(3), 856–880.

Piezunka, H., & Dahlander, L. (2019). Idea rejected, tie formed: Organizations' feedback on crowdsourced ideas. *Academy of Management Journal*, *62*(2), 503–530.

Plunkett, L. (2013). Steam Is 10 Today. Remember When It Sucked? *Kotaku US*, *9*, 12.

Poetz, M. K., & Schreier, M. (2012). The value of crowdsourcing: can users really compete with professionals in generating new product ideas? *Journal of Product Innovation Management*, *29*(2), 245–256.

Puranam, P., Alexy, O., & Reitzig, M. (2014). What's "new" about new forms of organizing? *Academy of Management Review*, *39*(2), 162–180.

Rietveld, J., & Eggers, J. P. (2018). Demand heterogeneity in platform markets: Implications for complementors. *Organization Science*, *29*(2), 304–322.

Rogers, E. M. (1962). *Diffusion of innovations*.

Schweisfurth, T. G. (2017). Comparing internal and external lead users as sources of innovation. *Research Policy*, *46*(1), 238–248.

Shah, S. (2005). Open beyond software. *Open Sources*, *2*, 339–360.

Shah, S. (2006). Motivation, governance, and the viability of hybrid forms in open source software development. *Management Science*, *52*(7), 1000–1014.

Shah, S., & Nagle, F. (2020). Why do user communities matter for strategy? *Strategic Management Review*.

Shah, S., & Tripsas, M. (2007). The accidental entrepreneur: the emergent and collective process of user entrepreneurship. *Strategic Entrepreneurship Journal*.

Sharkey, A. J., & Bromley, P. (2015). Can Ratings Have Indirect Effects? Evidence from the Organizational Response to Peers' Environmental Ratings. *American Sociological Review*, *39*(1), 63–91.

Shermon, A., & Moeen, M. (2020). Zooming in or zooming out: Entrants' product portfolios in the nascent drone industry. *Strategic Management Journal*.

Shontell, A. (2013). The Best Advice Airbnb CEO Brian Chesky Ever Received. *Business Insider*.

Singhal, A. (2001). Modern information retrieval: A brief overview. *IEEE Data Eng. Bull.*, *24*(4), 35–43.

Smith, S. W., & Shah, S. K. (2013). Do innovative users generate more useful insights? An analysis of corporate venture capital investments in the medical device industry. *Strategic Entrepreneurship Journal*, *7*(2), 151–167.

Sylvester, T. (2013). *Designing games: A guide to engineering experiences*. " O'Reilly Media, Inc."

Thomke, S. (2003). *Experimentation matters: unlocking the potential of new technologies for innovation*. Harvard Business Press.

Thomke, S., & Von Hippel, E. (2002). Customers as Innovators. *Harvard Business Review*, *80*(4), 74–81.

Toh, P. K., & Kim, T. (2013). Why put all your eggs in one basket? A competition-based view of

how technological uncertainty affects a firm's technological specialization. *Organization Science*, *24*(4), 1214–1236.
Vinokurova, N., & Kapoor, R. (2020). Converting Inventions into Innovations in Large Firms: How Inventors at Xerox Navigated the Innovation Process to Commercialize Their Ideas. *Strategic Management Journal*, 1–29.
Von Krogh, G., Haefliger, S., Spaeth, S., & Wallin, M. W. (2012). Carrots and rainbows: Motivation and social practice in open source software development. *MIS Quarterly*, 649–676.
Von Krogh, G., Spaeth, S., & Lakhani, K. R. (2003). Community, joining, and specialization in open source software innovation: a case study. *Research Policy*, *32*(7), 1217–1241.
Wang, D., Pahnke, E. C., & McDonald, R. M. (2021). The Past Is Prologue? Venture-Capital Syndicates' Collaborative Experience and Start-Up Exits. *Academy of Management Journal*, *ja*.
Welch, C. (2013). Steam Early Access lets gamers buy and play titles still in development. *Recuperado*, *6*(10), 2015.
Wijman, T. (2018). *Mobile revenues account for more than 50% of the global games market as it reaches $137.9 billion in 2018. Newzoo*.
Yau, H. K., & Tang, H. Y. H. (2018). Analyzing ecology of Internet marketing in small-and medium-sized enterprises (SMEs) with unsupervised-learning algorithm. *Journal of Marketing Analytics*, *6*(2), 53–61.
Yin, P.-L., Davis, J. P., & Muzyrya, Y. (2014). Entrepreneurial innovation: Killer apps in the iPhone ecosystem. *American Economic Review*, *104*(5), 255–259.
Zhao, E. Y., Ishihara, M., Jennings, P. D., & Lounsbury, M. (2018). Optimal distinctiveness in the console video game industry: An exemplar-based model of proto-category evolution. *Organization Science*, *29*(4), 588–611.
Zhu, F., & Iansiti, M. (2012). Entry into platform-based markets. *Strategic Management Journal*, *33*(1), 88–106.

Appendix

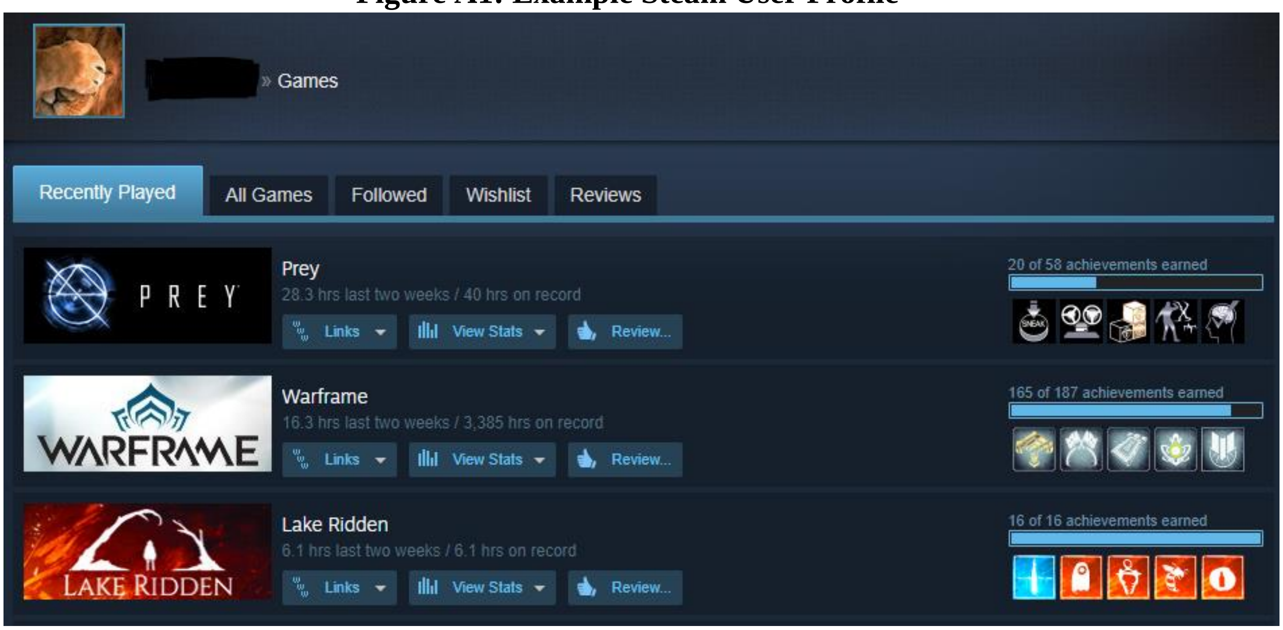
Appendix A: SteamSpy Data and Estimation Process

SteamSpy obtains its game ownership and usage statistics from Steam user profiles. Data on Steam user profiles are publicly accessible through Steam's web API. Each profile contains information on (1) which games the user owns, (2) whether the user has played each owned game in the past two weeks, and, if so, (3) the number of minutes the user spent playing that game in the past two weeks. See Figure A1 for an example Steam user profile. When data from all Steam user profiles is aggregated, it provides a cross-section of every purchase made on Steam, the number of active players for each game (in the past two weeks), and the average two-week playtime for each game.

SteamSpy repeats the process of querying and aggregating data from Steam user profiles daily to build a historical database of game ownership and usage statistics. Because the number of Steam user profiles exceeds 350 million (more than it is possible from SteamSpy to query in a single day), SteamSpy uses a web crawler to randomly sample approximately 500,000 Steam user profiles each day and generate estimates of statistics for each game (Galyonkin 2016).

To arrive at global estimates of the total number of owners and recent players for each game on Steam, SteamSpy first calculates the ratio of users in the random sample that own and have recently played each game. Next, these ratios are multiplied by the total number of Steam user profiles on the date of sampling (Orland 2015). Similarly, to arrive at global estimates of average and median playtime for each game, SteamSpy uses the average and median playtime of all users in the random sample. To reduce noise, SteamSpy calculates a 3-day rolling average for each estimate and rounds game ownership figures to the nearest 1000. Data is then made publicly available on the website steamspy.com, and has been used in several research studies (Bailey and Miyata 2019, Brunt et al. 2020, Lin et al. 2019).

Figure A1: Example Steam User Profile



Appendix B. Inverse Probability Treatment Weighting (IPTW)

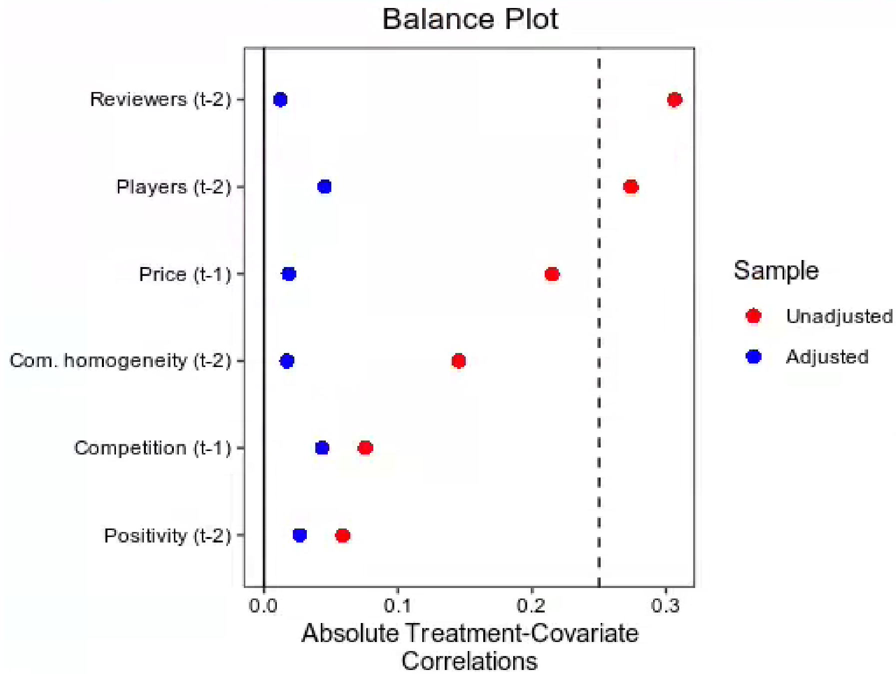
One endogeneity concern with our estimation strategy is self-selection into *Firm Responsiveness*. That is, certain features of developers or game projects may simultaneously predispose them to respond more to community feedback and to experience market growth. We used inverse probability treatment weights (IPTW) to address this concern (Azoulay et al. 2009, Wang et al. 2021). IPTW estimation generalizes propensity score weighting to treatments staggered over time. Thus, assuming that we can observe and measure the confounders, IPTW allows us to estimate treatment effects even in the presence of time-varying confounders. This is done by calculating weights for each observation, that assigns higher weights to observations that were less likely to be treated. This effectively creates a pseudo-population whose confounders no longer predict selection into treatment, but where the causal relationship between the treatment and outcome is the same as the original population (Azoulay et al. 2009). Formally, weights are calculated as the probability distribution of treatment divided by the probability distribution of the treatment variable explained by the confounders:

$$W_{it} = \prod_{t=1}^{t} \frac{f_X(X; \mu_1, \sigma_1^2)}{f_{X|C}(X|C = c; \mu_2, \sigma_2^2)}$$

Where X stands for the continuous treatment variable, C stands for the confounders, and the function f is notation for a probability distribution function with mean μ and variance σ^2.

To adjust for observable confounders, we included the following measures as proxies for market demand and game quality: $Reviewers_{t-2}$, $Players_{t-2}$, $Price_{t-1}$, *Community homogeneity*$_{t-2}$, $Competition_{t-1}$, $Positivity_{t-2}$. We report variable balances in Figure B1. Weights were truncated at 0.1 to achieve better balance. After IPTW adjustment, all confounders' correlations were less than the absolute standardized mean difference of 0.25 recommended by Rubin (2001, p. 174) and Stuart (2010, p. 11).

Figure B1. Covariate Balance from Inverse Propensity Weighted Treatment



Notes: these figures compare the covariate balancing of treatment (*Firm Responseiveness*$_{it-1}$) for the unadjusted sample (red dots) and the adjusted sample (blue dots). The adjusted sample represents the balance after applying the inverse propensity weighting procedure.

Appendix References

Azoulay P, Ding W, Stuart T (2009) The impact of academic patenting on the rate, quality and direction of (public) research output. *J. Ind. Econ.* 57(4):637–676.

Bailey E, Miyata K (2019) Improving video game project scope decisions with data: An analysis of achievements and game completion rates. *Entertain. Comput.* 31:100299.

Brunt CS, King AS, King JT (2020) The influence of user-generated content on video game demand. *J. Cult. Econ.* 44(1):35–56.

Galyonkin S (2016) Steam sales in 2015. *Retrieved March* 31:2017.

Lin D, Bezemer CP, Zou Y, Hassan AE (2019) An empirical study of game reviews on the Steam platform. *Empir. Softw. Eng.* 24(1):170–207.

Orland K (2015) Introducing Steam Gauge: Ars reveals Steam's most popular games. ars technica.

Wang D, Pahnke EC, McDonald RM (2021) The Past Is Prologue? Venture-Capital Syndicates' Collaborative Experience and Start-Up Exits. *Acad. Manag. J.* (ja).