Cohering with the Crowd: How Audiences Shape the Quasi-Scientific Process of Entrepreneurship

By

Simon C.A.T. Friis

B.Sc. Asian Business, Language, and Culture,
Copenhagen Business School, 2014

S.M. Management Research,
MIT Sloan School of Management, 2018

SUBMITTED TO THE DEPARTMENT OF MANAGEMENT IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY IN MANAGEMENT

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

MAY 2022

© 2022 Massachusetts Institute of Technology. All rights reserved.

Signature of Author: ____
Department of Management
May 6, 2022

Certified by: ____
Ezra W. Zuckerman Sivan
Alvin J. Siteman (1948) Professor of Entrepreneurship and Strategy
Professor, Technological Innovation, Entrepreneurship, and Strategic Management and Work and Organization Studies
Thesis Supervisor

Accepted by: ____
Catherine Tucker
Sloan Distinguished Professor of Management
Professor, Marketing
Faculty Chair, MIT Sloan PhD Program

Cohering with the Crowd: How Audiences Shape the
Quasi-Scientific Process of Entrepreneurship

By

Simon C.A.T. Friis

Submitted to the Department of Management on May 6, 2022,
in Partial Fulfillment of the Requirements for the Degree of
Doctor of Philosophy in Management

ABSTRACT

Important aspects of entrepreneurship can be usefully understood as a quasi-scientific process in which entrepreneurs develop theories of value and test those theories through experimentation. Unlike academic scientists, however, entrepreneurs often develop and test theories in collaboration with an audience. The impact of audiences on the quasi-scientific process is brought into sharp relief on the livestreaming platform *Twitch.tv*, where entrepreneurs compete in a cultural market for the scarce attention of viewers.

The first essay examines how theories of value constrain strategic choice and valuation. I ask: why are some combinations of product categories more appealing to audiences than others? A prominent line of work, drawing on prototype theory, posits a universal penalty for category-spanning offerings. I clarify the limitations of this approach, focusing in particular on its inability to explain change. I introduce theoretical coherence (the extent to which a combination of product categories coheres with a theory of value) as an alternative standard for understanding the appeal of categorical combinations. I develop and validate an empirical framework that uses word embedding models to study theoretical coherence and find that theoretical coherence is able to explain the appeal of product-category combinations not easily addressed by prototype theory.

The second essay examines why successful experimentation requires the effective collaboration of audiences and how this in turn limits the strategic opportunities of entrepreneurs. Experimentation is traditionally thought to improve entrepreneurial outcomes because it avoids costly commitment and allows entrepreneurs to pivot to more attractive product markets. I develop and test a theory that recognizes the costs experiments impose on audiences. My theory implies that successful experimentation involves a tradeoff between two types of commitment. On the one hand, an entrepreneur can invest in developing a better prototype, thereby increasing the audience's willingness to test the prototype. On the other hand, an entrepreneur can focus on developing their relationship with their audience, thereby increasing the audience's tolerance for crude prototypes. I find that *Twitch* streamers who invest more in developing relationships with their audience experience fewer penalties from experimentation but get trapped in less attractive product markets.

Thesis Supervisor: Ezra W. Zuckerman Sivan
Title: Alvin J. Siteman (1948) Professor of Entrepreneurship and Strategy, Professor, Technological Innovation, Entrepreneurship, and Strategic Management and Work and Organization Studies

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

When I first set foot in Cambridge in 2014, I was eager to face the challenges of the PhD head on. Perhaps it was naïveté or perhaps it was hubris, but the challenge turned out to be far greater than I anticipated. Although everyone takes a different path, it is safe to say that no one does it alone. To use a trite but apt metaphor: it takes a village to raise a person. I would not have been able to complete this journey without the help of many people. Some had to bear the brunt of that support more than others, but the final outcome wouldn't have been possible without the concerted efforts of so many others.

More than anyone, Ezra Zuckerman Sivan has molded me into the scholar I am today. When I took Ezra's course on "Identity and Action" in my first year at Sloan, it seemed like no matter what I would say, he would probe my remarks and push me to clarify what I meant. The short-term effect of this was that I was too terrified to say anything for the first few weeks, as it became obvious to me that I really had no idea what I was saying! The long-term was to instill in me a taste for good research and especially a taste for good theory. More importantly though, working with Ezra has changed the way I view the world. Ezra's effect on his students is in no small part due to the remarkable dedication he has for them. Despite taking on a deanship, he would always make time for me. Any draft that I would send him would be returned with remarkable speed, commenting on everything from the theoretical contribution to comma placement.

It wasn't until later in my graduate school career that Ray Reagans began advising me. But in that short time, he has provided invaluable mentorship and played a pivotal role in my scholarly development. My only regret in working with him is that I have not been able to digest and incorporate more of his feedback. He is the kind of advisor that can comfortably go from highlighting connections to seminal theoretical work from the 1940s to having an in-depth discussion about the latest natural language processing techniques, all while showing deep concern for his students' wellbeing. Ray had also taken on a deanship, yet continued to make time for me.

I have been truly fortunate to have two advisors so devoted to passing on the craft of research to the next generation. Their care for their students is evident in the depth and seriousness with which they engage in our work. Their high standards in research are matched only by their deep personal regard for their students. I can only hope to strive for their standards in my mentorship of others.

Many other faculty have contributed to my scholarly development. In addition to having had the pleasure of working closely with them on projects and learning from them, Emilio Castilla and Christian Catalini were instrumental in seeing me on to the next stage of my career. One of the things that makes Sloan such a special place is the willingness of faculty across groups to support students. I have had the privilege of learning from Abdullah Almaatouq, Sinan Aral, Pierre Azoulay, Lori Breslow, John Carroll, Jared Curhan, Barbara Dyer, Roberto Fernandez, Neal Hartman, Olenka Kacperczyk, Miro Kazakoff, Erin Kelley, Kate Kellogg, Tom Kochan, Danielle Li, Jackson Lu, Fionna Murray, Wanda Orlikowski, Erin Scott, Ben Shields, Susan Silbey, Scott Stern, Basima Tewfik, Catherine Turco, Inez Freein von Weitershausen, Nate Wilmers, and JoAnne Yates.

During my grad school admission interviews, I was often asked "what made you apply to this program?" Ironically, I doubt the organizational scholars asking the question would accept path dependence or serendipity as good answers. But my path to grad school, and specifically to

Sloan, was very much an act of serendipity. The first faculty member I asked to supervise my undergraduate thesis turned me down and suggested I ask Louise Mors Christensen instead. I could not have been luckier. She took me under her wing, gave me invaluable research experience, and saw me through every step of the application process. I am also grateful to Christian Geisler Asmussen and Yang Jiang for their support.

My fellow PhD students have sat through countless presentations, indulged half-baked ideas, and provided emotional support. I am fortunate to not only have you as my colleagues, but also as true friends. Among those who I have known the longest are Brittany Bond, Avinash Collis, Vanessa Conzon, Caroline Fry, Carolyn Fu, Rebecca Karp, Mahreen Khan, Minjae Kim, James Riley, Melissa Staha, Michael Wahlen, Duanyi Yang, and Heather Yang. I've also had the pleasure of getting to know your partners, Nick, Manuella, Eric, Mark, Charlie, Megan, Yu, and Collin.

So many other peers have enriched my Sloan experience: Tristan Botelho, Donato Cutolo, Paramveer Dhillon, Erik Duhaime, Leonardo Elias, Dan Fehder, Ari Galper, Laurel Grassin-Drake, Wesley Greenblatt, Jorge Guzman, Jaekyung Ha, Mahdi Hashemian, Dave Holtz, James Houghton, Summer Jackson, Arvind Karunakaran, Raquel Kessinger, Danny Kim, Hyejun Kim, Soomi Kim, Will Kimball, Joshua Krieger, Alex Kowalski, Tatiana Labuzova, Jose Lopez, Audrey Mang, Fernando Martins, Claire McKenna, Maarten Meeuwis, James Mellody, Arrow Minster, Zanele Munyikwa, Jenna Myers, Christina Nguyen, Ethan Poskanzer, Hye Jin Rho, Christine Riordan, Eppa Rixey, Dan Rock, MacKenzie Scott, Bradley Turner, Hagay Volvosky, George Ward, Jane Wu, Alan Zhang, Victoria Zhang, and Samantha Zyontz.

The administrators and staff at Sloan made everything work behind the scenes. I am grateful to Ellen Baum, Colette Boudreau, Patty Charest, Natalia Kalas, Sarah Massey, Hillary Ross, and Davin Schnappauf, and Rajiv Shridhar for their support. On my frequent trips to visit Ezra in E-60, I always looked forward to seeing Gilly Parker, who never failed to brighten my day. I had the pleasure of working with Sheila Canalle and Debra Luchanin throughout most of my graduate career. It was also a joy working with Katy Bertman, Mahren Reisch, Abby Berenson, Chan Phanh, and Jessica Sumney. I had the pleasure of getting to know Jenn Alton, Shallon Silvestrone, and Channa Yem outside of Sloan.

Graduate school is in some ways a greedy institution. It is easy to get trapped in an academic bubble and lose sight of the outside world, especially in Cambridge. Harald Gjerde, Tony and Yuko Hanada, Lars Frølund and Karine Blandel, Christian Madsen, and Che Yeun provided an escape from that bubble and invaluable friendship.

Another stroke of luck was that graduate school would bring me closer to my extended family. Amir, Angela, Nicole, Brendan, and Zach always included me in their holiday festivities and gave me a sense of family and roots here in the States. One of the great joys of attending graduate school in the US is that I was able to grow closer to them, as well as to see Massoud and Said, even if just a few times. Scarlett, Darrin, and Tate sent much love and support from the UK. My grandmother continues to be the kindest and most loving person I know. A quick chat with her never fails to cheer me up.

Mette and Børge have been much more than aunt and uncle to me. They have played a tremendous role in my life and no less my PhD. Words cannot express all they have done for me and my appreciation of them.

Even harder to put into words is my gratitude for my parents, who have provided a lifetime of unconditional love and unwavering support. Their presence is felt in everything I do. I had hoped that I would finally be able to outsmart them with a PhD, but to my great frustration they remain far wiser than I.

Finally, I am forever grateful to Helen, my best friend, who stalwartly stood by my side through all the ups and downs of the last four years. She kept me going no matter what. Her humor, strength, intellect, and curiosity inspire me every day.

Thank you all.

CHAPTER 1

INTRODUCTION

In 2015, Tyler Blevins joined a niche community of “streamers” on the platform *Twitch*. Tyler and other streamers would use *Twitch* to broadcast live video of themselves playing videogames to anyone who cared to watch. To say that Tyler is an avid gamer is an understatement. In 2017, he spent 3,203 hours livestreaming his gameplay on *Twitch,* an average of nearly nine hours per day (SullyGnome, n.d.).

When the videogame *Fortnite* was released in 2017, Tyler was one of its early adopters and would play it on his *Twitch* channel. In only a few months after its release, *Fortnite* became a global sensation, amassing 30 million players. Fortunately for Tyler, he was in the right place at the right time. Tyler—who is better known on *Twitch* as “Ninja”—and his signature electric-blue hair would become the face not just of *Fortnite* but also livestreaming more generally. As the meteoric rise of *Fortnite* continued unabated into 2018, Ninja would regularly have over one hundred thousand people watching him at any given time. Ninja had become a household name and videogame livestreaming, once a niche activity, had been thrust into the mainstream. Ninja’s celebrity status was lucrative; he reported earning at least $500,000 per month (Heitner 2018). Contrary to the advice of sensible parents everywhere who told their children there was no future in videogames, Ninja demonstrated to a whole generation that videogames and livestreaming could in fact be a viable career.

Although the growing popularity of Ninja and *Fortnite* showed no signs of slowing down, Ninja faced a looming strategic dilemma: videogames are heavy in fads and fashion cycles. Neither Ninja nor *Fortnite’s* popularity was guaranteed; they could be forgotten just as quickly as they had been thrust into the spotlight. If Ninja’s viewership is tied to the popularity of *Fortnite,* what should he do when viewers move on to the next big game? For streamers in Ninja’s position, who wish to retain their hard-earned viewers but also continue to build their audience, they must be strategic about how they build their game repertoires. One common strategy for reducing risks in turbulent environments is to diversify (Freeman and Hannan 1983). In Ninja's case, this would mean adding additional games to his repertoire that he could play for his viewers, much like an investor who constructs a portfolio of assets to limit their risk.

But such a strategy is challenging on *Twitch,* where audiences are notoriously fickle. Given the thousands of other streamers *Twitch*—all of whom can be watched for free—viewers have essentially no switching costs. Ninja reported losing 40,000 subscribers when he took a two-day break from streaming to compete in a charity gaming tournament (Grayson 2018), equivalent to at least $100,000 of monthly income. Just as viewers place strict demands on the availability of streamers, they also place strict demands on the games they are willing to watch. Many of Ninja's viewers are strictly interested in watching him play *Fortnite* and would abandon him if he were to play other games.

At the same time, it is reasonable to expect that Ninja's audience is going to be more accepting of some games than others. Careful selection of games can thus mitigate some of the penalties he might otherwise face. At least since Rumelt (1974), strategy scholars have thought related diversification to be better than unrelated diversification. Suggesting that Ninja expand his repertoire by adopting games similar to *Fortnite* seems intuitive, but closer examination reveals deeper challenges with this approach. In particular, what does it mean for a game to be "related" to *Fortnite*? While strategy scholars have long debated the question of what constitutes relatedness, one prominent idea is that firms should diversify into areas that allow them to exploit and develop their core competencies (Markides and Williamson 1994). For example, as a former professional *Halo* player and decorated champion in *Fortnite,* Ninja certainly has particular gaming skills that he could apply to great effect in other games of similar genre. If all

Ninja's viewers cared about was observing the highest possible levels of gameplay, the answer might be straightforward: Ninja should choose the game with the greatest skill carryover (a less helpful implication of this view is that he should specialize in one game). As the following chapters discuss in detail, even though audiences on *Twitch* can and do care about skill in many instances, a streamer's personality and their relationship to their viewers can be even more important. This begins to raise the question: how does someone like Ninja know what his core competencies are? And insofar as one of his core competencies relates to his personality, what exactly does this imply for the kinds of games he should pursue?[1]

A less obvious limitation of this approach is that the logic behind what kinds of related diversification are "best" traditionally focuses on supply-side factors, like skills and resources. But compared to the knowledge and capital-intensive industries traditionally studied in strategy research, producers on *Twitch* face relatively few supply-side constraints on strategic choice and opportunity. Entry and exit costs are negligible: the platform is free to join and the software and hardware required for streaming is widely available. Because most streamers are individuals, they face strict limits in their ability to scale production; they can only stream as much as the number of hours in a day. And specifically in terms of constructing repertoires, streamers have access to the same games and can thus readily construct the same repertoires.[2]

The more salient constraints on strategic choice and opportunity that *Twitch* streamers face are on the demand side. In other words, the primary constraints are imposed by the audience. Demand-side constraints are often salient in cultural markets, such as markets for art,

---

[1] These challenges are not inconsistent with the need for related diversification or the importance of core competencies. A producer's identity can be considered a strategic asset and a core competency of producers in cultural markets, especially in today's social media ecosystem, is their ability to craft appealing identities. However, this explanation borders on tautology and does not readily explain why one combination of games would be better than another.

[2] There are, of course, exceptions. In particular, some high-profile streamers receive early access to games. However, early-access game play constitutes a relatively minor proportion of these streamers' gameplay.

where audiences are sensitive to the identities of producers and place limits on the range of identities they are willing to accept. Research by economic sociologists and organization theorists has long documented that audiences will penalize or ignore producers who present incoherent or illegitimate identities (for reviews of this literature, see Negro, Koçak, and Hsu 2010; Vergne and Wry 2013). For example, a large brewing company like Anheuser-Busch may have distinct competencies that would give them a competitive advantage in producing craft beer. Yet, numerous studies document how craft beer enthusiasts penalize beers that are associated with corporate brewers for being inauthentic (Carroll and Swaminathan 2000; Frake 2017).

When viewed from the demand side, the primary strategic question that a streamer like Ninja must answer when putting together a repertoire of games is: what combinations of games will my audience accept and why? This dissertation takes this question as its point of departure and uses the context of *Twitch* to make progress in our understanding of the demand-side factors that constrain strategic choice and opportunity. As will become clearer, considering the "practical" strategic challenges that streamers like Ninja must deal with reveals deeper theoretical and empirical issues, which are the primary focus of this dissertation.

ENTREPRENEURS AS QUASI-SCIENTISTS

To understand how entrepreneurs like Ninja navigate the uncertainty that makes these strategic decisions challenging, a useful metaphor is that of the "entrepreneur as quasi-scientist" (Felin et al. 2020; Felin and Zenger 2009, 2017; Camuffo et al. 2020). Much like academic scientists, entrepreneurs develop theories and test them through experimentation. While the metaphor is useful for understanding positive and normative aspects of entrepreneurial strategy, it is also limited.

Entrepreneurs of course differ from scientists in important ways. Unlike a physicist and the atoms they may study, entrepreneurs study audiences. The physicist and the atom have a certain mutual disregard for one another. Passionate disinterestedness is the hallmark of the consummate scientist. The atom has even less regard for the physicist. It does not care which physicist studies them, nor does it mind if the physicist wants to look at other atoms. By contrast, entrepreneurs and audiences often take a personal interest in one another for their fates are intertwined. In cultural markets, for example, the identities of producers and audiences are often fused. In the context of science, close, personal relationships between the scientist and the objects they study represent threats to validity. By contrast, in the context of entrepreneurship, these relationships may represent opportunities for joint value creation. Entrepreneurs thus have ample reason not to treat audiences merely as subjects to be studied, but as collaborators in the quasi-scientific process. I revisit this metaphor in more depth in the final chapter.

In this dissertation, I examine the role that audiences play in enabling and constraining entrepreneurial theorization and experimentation. The rest of this dissertation is structured as follows. Chapter 2 provides a broad introduction to the primary setting, *Twitch.* Chapter 3 describes features of my *Twitch* data that are common to subsequent chapters.

Chapter 4 examines how theories of value constrain strategic choice and valuation. I ask: why are some combinations of categories are more viable than others? One prominent approach to this question emphasizes the importance of prototypicality, but this approach is problematic because it unable to explain why the same combination may be viable for one audience but not another and why the viability of combinations changes over time. To overcome these limitations, I develop and test an explanation based on theoretical coherence—the extent to which a combination of categories coheres with a theory of value. Using word embedding models and methods from computational social science, I develop and validate an empirical framework for analyzing theoretical coherence. By examining two communities focused on competitive videogaming—eSports and speedrunning—I show how nuanced differences in their theories of skill lead to different constraints and valuation of their game repertoires.

Chapter 5 examines the role of audiences in shaping the dynamics of entrepreneurial experimentation. I ask: why are some entrepreneurs more successful at experimentation than others? Experimentation is traditionally thought to improve entrepreneurial outcomes because it avoids costly commitment. I develop and test a theory of entrepreneurial experimentation that recognizes the costs that experiments impose on audiences. In contrast to the traditional view, my theory implies that successful experimentation in fact requires a tradeoff between two types of commitment. An entrepreneur can either pre-commit to their audience to increase their tolerance of crude prototypes, or the entrepreneur can pre-commit to developing a better prototype that audiences are more willing to tolerate.

CHAPTER 2

TWITCH.TV: A MARKETPLACE OF COMMUNITIES

THE EMERGENCE OF LIVESTREAMING: FROM LIFECASTING TO ESPORTS

On March 19, 2007, Justin Kan started an experiment that would last only eight months, but would fundamentally change the modern social media ecosystem. He attached a webcam to a baseball cap and began broadcasting his entire life to the Internet (Beale 2007). By carrying a laptop connected to a cell phone in a backpack, he was able to broadcast 24/7 from any location, stopping only for bathroom and bathing breaks. The live video stream was freely accessible to the public through his website *Justin.tv*, where viewers could chat with him while they watched. As *Justin.tv* continued to grow in popularity, the site expanded in October 2007 to allow others to create livestreams of their own. Building on Justin's "lifecasting," the original plan was to focus on IRL (in real life) content and create a kind of 24/7 reality TV (Rice 2012). Early livestreams—or simply *streams*—on *Justin.tv* featured content in a variety of categories, including Music, Animals, and "Divas and Dudes" (Rice 2012).

Given the early ambitions to create a reality TV network, a seemingly unlikely contender emerged, quickly dwarfing the rest of the categories: gaming. Rather than play videogames themselves, viewers could instead watch others play videogames. Gaming streams ranged from people casually playing *Wrestlemania* with friends to professional eSports players competing in *StarCraft* tournaments.

In hindsight, the stage was ripe for videogame livestreaming. Videogames, once a niche activity, were becoming mainstream. In the United States, retail sales of video games nearly doubled over a three-year period, growing from 12.3 billion USD in 2005 to 23.1 billion USD in 2008 (Statista 2015). Interest in eSports was also growing rapidly, although reliable statistics for that time period are limited. These communities had deep roots in the Internet, and in the case of eSports, previously limited in their ability to reach fans. Technology played a pivotal role as well. On the broadcasting side, *Justin.tv* provided improved video compression, allowing home users to broadcast high-quality video more efficiently than before. At the same time, high-speed internet access was becoming more common, giving people the necessary bandwidth to both broadcast and watch live video. All the while, the costs of streaming equipment such as webcams and computers were falling, lowering the barriers to entry for potential streamers.

In June 2011, the gaming channel was spun out to its own separate website, *Twitch.tv*, often referred to simply as *Twitch.*[3] In the first quarter of 2022, *Twitch* boasted 10.9 million active streamers and an average of 2.8 million concurrent viewers (May 2022). In total, users watched over 6.1 billion hours of content. From its humble beginnings, videogame livestreaming is now unquestionably in the mainstream.

The livestreaming industry

A number of platforms dedicated to videogame livestreaming have since emerged. Major competitors to *Twitch* include *YouTube Gaming*, *Facebook Gaming*, and *Microsoft*'s now defunct *Mixer*. *Twitch*, which was acquired by Amazon in 2014 for 970 million US dollars (Gittleson 2014), has consistently been the market leader in this industry. As seen in figures Figure 2-1 and Figure 2-2, *Twitch* is by far the most popular videogame livestreaming platform with regard to both streamers and viewers. In the fourth quarter of 2020, hours watched on *Twitch* was nearly twice that of *YouTube Gaming* and *Facebook Gaming* combined and nearly ten times the number of hours streamed.

[3] The name Twitch refers to "twitch gameplay," which is gameplay that requires ultrafast reflexes and fine-grained motor control (*Twitch gameplay, Wikipedia*).

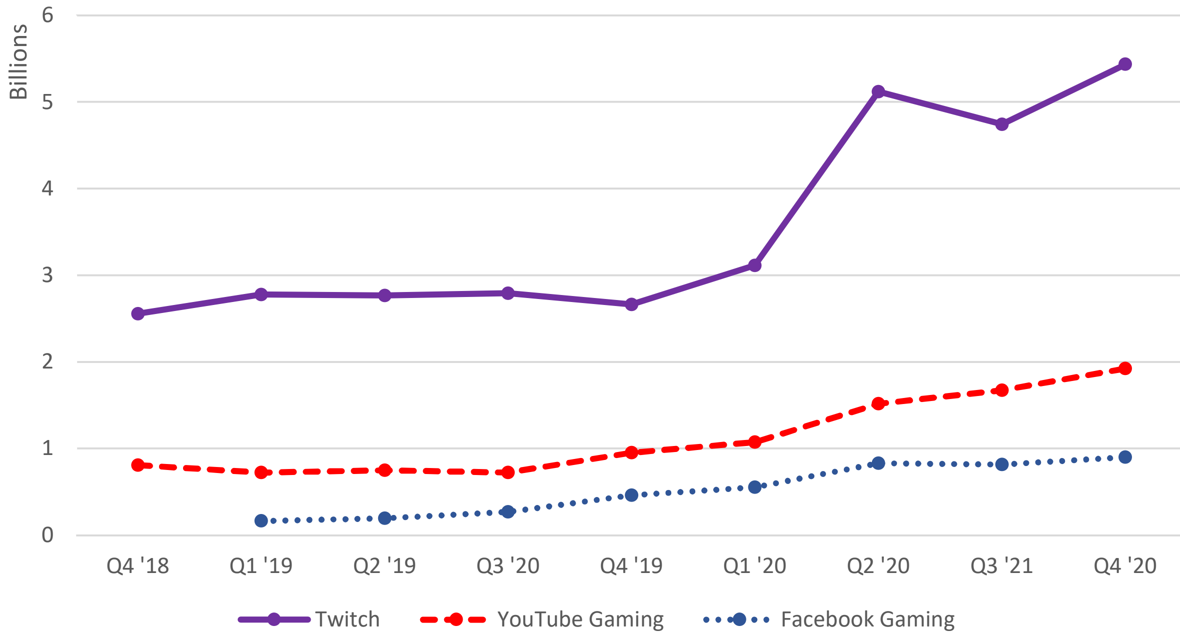

Figure 2-1. Total hours watched across major videogame livestreaming platforms. Source: Adapted from May (2021b). Note: Data not available for Facebook Gaming in Q4 '18.

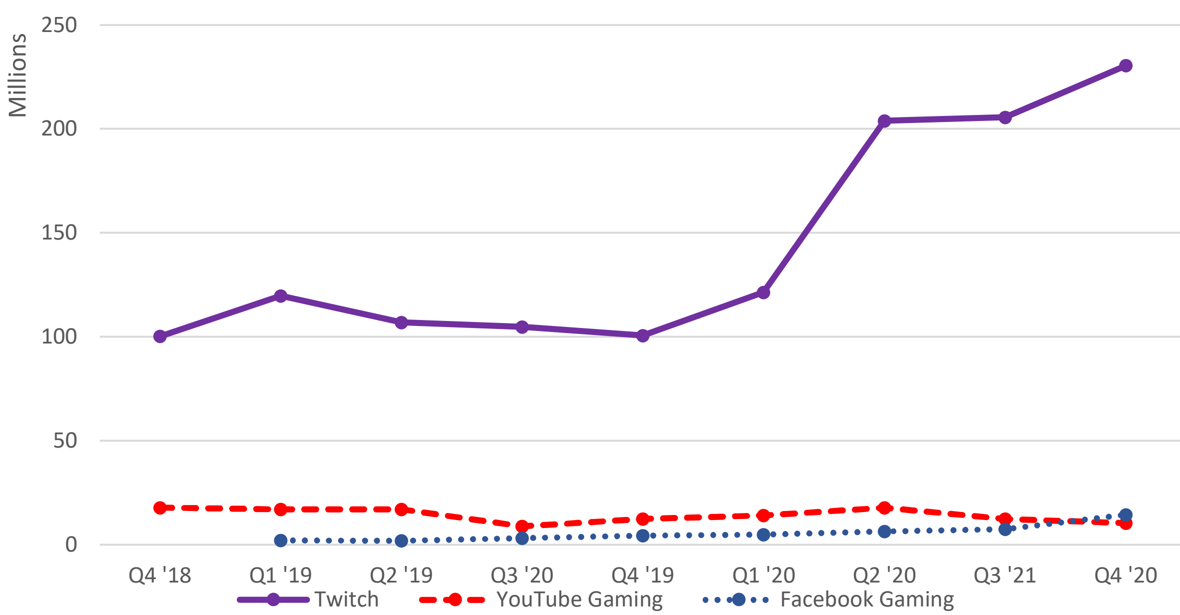

Figure 2-2. Total hours streamed across major videogame livestreaming platforms. Source: Adapted from May (2021b). Note: Data not available for Facebook Gaming in Q4 '18.

The livestreaming experience

Livestreaming is unique among social media in that it occurs in real-time and is highly interactive. Figure 2-3 shows a typical *Twitch* stream, which includes a public chat box to the right of the video. Viewers chat with each other as well as the streamer, who typically responds verbally. Streamers and their viewers often refer to themselves as a "community." In smaller streams, streamers and regular viewers even come to know each other on a personal basis. These interactions can extend to in-person meetups at events like the annual TwitchCon (Taylor 2018).

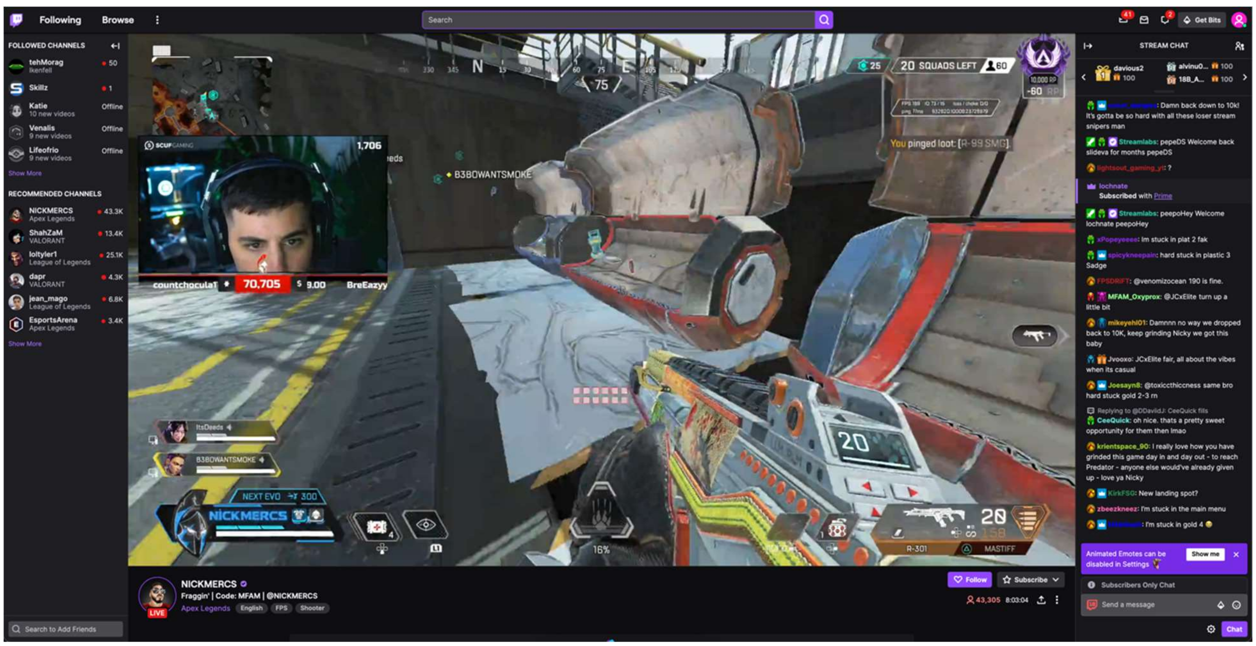

Figure 2-3. A typical *Twitch* stream. The game takes up the majority of the screen and the streamer is shown in a picture-in-picture frame. Viewers can chat with each other and the streamer using the chatbox on the right.

Although Twitch was originally created for videogame livestreaming, the site has expanded to include a variety of content, ranging from coding to fishing to political commentary. A particularly noteworthy category "Just Chatting," which has been the most popular category since 2020 (May 2022). As the name implies, Just Chatting features streamers primarily chatting and interacting with their audiences and not necessarily engaged in a particular activity. However, the content within this category is also highly varied. For example, "mukbang" streams are a type of social eating, where the streamer eats (often extravagant) meals in front of the camera while chatting with the audience. Other content in the Just Chatting category is borderline pornographic, such as controversial "hot tub streams." (*Twitch Blog* 2021).[4]

However, gaming content still comprises the vast majority of content on Twitch, both in terms of the number of categories and activity. On Twitch, each game has its own dedicated category, resulting in more than 70 thousand categories dedicated to video games. For example, games in the same franchise such as *Super Mario World* and *Super Mario 64* are distinct categories. By contrast, there are only 15 non-video game categories[5].

Stream categories are important to both streamers and viewers. One of the only ways of browsing streams on Twitch is by category. Figure 2-4 shows how users navigate the site. When a user clicks on a category, they are shown a list of all the streams in that category. Streamers choose how their stream should be categorized. In most cases the categorization is straightforward: it corresponds to the videogame that is being played. Only one category can be applied to a stream at a time, although the streamer can change categories within a streaming session. This would be the case, if, for example, the streamer switched from playing *Fortnite* to playing *Call of Duty*. Categories are set and maintained by Twitch, so streamers cannot create new categories.

[4] *Twitch*'s rules of conduct prohibit nudity and pornographic material. This content poses a number of questions of broader sociological interest relating to gendered labor on digital platforms and moderation of digital platforms. These questions are taking on increasing importance as more work and life moves online.

[5] The 15 IRL (In Real Life) categories are: Just Chatting, Music, Creative, Sports, Art, ASMR, Talk Shows & Podcasts, Science & Technology, Makers & Crafting, Travel & Outdoors, Food & Drink, Politics, Fitness & Health, Special Events, Bex`auty & Body Art.

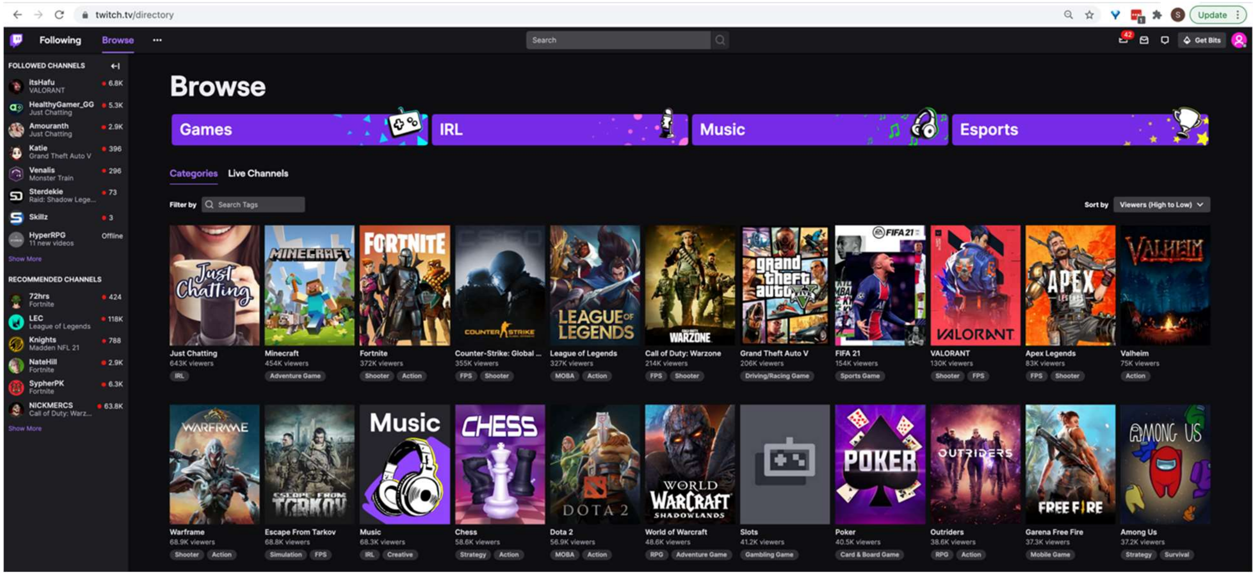

Figure 2-4. The *Twitch* directory. Each game is its own category.

In addition to categories, streamers can apply tags to their streams. Tags are intended to provide additional context about the stream and are largely orthogonal to the categories. If categories represent what the streamer is doing, stream tags represent how the streamer is doing it. Streamers can choose to apply up to five tags to their stream. Similar to categories, the list of available tags is set and maintained by Twitch.

Professionalization and monetization

Most streamers are part-time hobbyists, but streaming is increasingly a viable full-time occupation. Although *Twitch* is free for both streamers and viewers, the platform enables streamers to monetize their content and earn revenue in several ways. One important source of revenue is subscriptions. *Twitch* offers three tiers of subscriptions at $4.99, $9.99, and $24.99 per month. The streamer typically receives 50% of this amount while *Twitch* retains the remainder (highly popular streamers may negotiate different terms). Subscriptions are primarily seen as a donation to the streamer and provide few tangible benefits. Common benefits include a subscriber badge and subscriber-only emojis (small pictures) that can be used in chat. Some highly active streams limit the chat to subscribers only and some Partners (explained below) have the option to disable ads for subscribers. As explained in the affiliate and partner status section below, these monetization options are available only to those streamers who meet certain performance criteria and are awarded Affiliate or Partner status. In addition to the monetization options provided by Twitch, streamers are able to generate revenue through a variety of other sources, such as ad revenue sharing, donations, corporate sponsorships, and referral codes.

Given that monetization is closely linked to audience size, building an audience is central to success on *Twitch*. Like many other content creators, streamers engage in intimate, relational work to build personal connections with their viewers, who may in turn choose to financially support them (Baym 2018; Taylor 2018). Like other bifurcated cultural markets, streamers at the top, like the popular player of the videogame *Fortnite* named “Ninja,” have become household names and earn millions of dollars. Meanwhile, most streamers struggle to attract even as few as ten viewers. At any given time, thousands of people are broadcasting to an audience of zero in hope that viewers will soon arrive (Hernandez 2018). Given the large number of streamers, the low barriers to entry, and negligible switching costs for viewers, streamers face great competitive pressure to build and maintain their audiences. However, streamers do not necessarily seek to maximize their audience size. Some streamers and viewers prefer smaller communities because these provide greater opportunities for close interpersonal interaction. Nevertheless, an important feature of this setting is that all streamers can reasonably be expected to be sensitive to viewership, regardless of whether their primary concern is the quantity or quality of viewership.

Affiliate and Partner status

After meeting certain performance requirements, *Twitch* streamers become eligible for Affiliate and Partner status (see *Twitch* n.d.-a, n.d.-b, n.d.-c). Achieving these statuses marks a milestone in the development of a streamer's channel. Importantly, the monetization options provided by *Twitch* are only available to Affiliates and Partners.

The minimum qualification criteria for Affiliate and Partner status are well defined by *Twitch*. To qualify for Affiliate status, streamers must not yet be Partners, have at least 50 followers and over the prior 30 days have at least 500 total minutes broadcast, 7 unique broadcast days, and an average of 3 or more concurrent viewers. Once these requirements are met, a button will appear that when clicked grants the streamer Affiliate status.

The eligibility requirements for Partner status are stricter than for Affiliate. Over the past 30 days, streamers must have at least 25 hours broadcast, 12 unique broadcast days, and an average of 75 or more concurrent viewers. In contrast to Affiliate status, meeting the Partner requirements does not guarantee Partner status. Meeting these requirements makes the streamer eligible to apply for Partner status. Applications are manually reviewed by Twitch staff. Twitch Partners are expected to adhere to stricter community and content guidelines and act as "role models to the community" (*Twitch* n.d.-c).

CHAPTER 3 DATA

This chapter provides a brief overview of the *Twitch* data, which is used in subsequent chapters. Additional details about data sources are provided where relevant.

The primary data used in this dissertation is a panel dataset of all *Twitch* streams that occurred between October 2019 and February 2022. All active livestreams were observed at 15-minute intervals during this period. As such, streams lasting longer than 15 minutes are observed at multiple points in time. These data thus afford an unusually comprehensive view of the dynamics of an entire platform at high temporal resolution. Descriptive statistics for the data at various levels of aggregation are shown in Table 3-1.

Two variables are of particular interest: the category the stream is in and the number of concurrent viewers. As discussed in the previous chapter, nearly all of the categories on *Twitch* correspond to individual games. Thus, the games *Diablo* and *Diablo II* are distinct categories. The number of concurrent viewers refers to the number of viewers at a given point in time. By observing streams at 15-minute intervals, I am thus not only able to track the evolution of streamers' game repertoires, but the near-instantaneous changes in viewership associated with changes in the game repertoire.

The 15-minute interval strikes a balance between data quality and computational feasibility. Because streamers can change categories within a single streaming session and changes in viewership are likely to occur immediately after changing categories, it is important to make the collection interval as small as possible to capture these granular changes. At the same time, it is unlikely that a streamer will spend less than 15 minutes in a particular category as changing categories so frequently would be disruptive. On average, streamers in the dataset will continuously stream in one category for 2.3 hours. While an interval shorter than 15 minutes would lead to a marginal improvement in accuracy, given the large number of streams that occur on *Twitch*, this would lead to significant increases in computational complexity.

Unit of analysis: aggregation to game-sessions

The primary outcome of interest in subsequent chapters is the change in a streamer's viewership associated with playing a particular game. Thus, I aggregate the raw data to what I call the game-session, which I define as an uninterrupted period streaming in one category. A single streaming session can thus be associated with one or more game-sessions. Recall that categories can change within a streaming session and across streams, but that each observation in the raw data is associated with at most one category.

Figure 3-1 depicts how category sessions are nested within streaming sessions. In this example, the streamer has two separate streaming sessions and a total of five category sessions. In the first stream session, the streamer logs on, starts streaming in the Just Chatting category, moves to playing *Fortnite*, goes back to Just Chatting, and then logs off. The next time the streamer logs on, they play *Call of Duty* and *Fortnite* before logging off.

Figure 3-1. Depiction of category session. Category sessions are an uninterrupted spell streaming one category and are nested within streaming sessions. Stream session 1 consists of three category sessions and stream session 2 consists of two category sessions.

There are two potential data quality issues with categorization: (1) Twitch does not require streamers to set a category, and (2) it would be cause for concern if streamers were actively miscategorizing their streams, perhaps to deceive viewers. These issues might raise concerns about self-selection into categorization (vs. non-categorization) or miscategorization. However, neither issue is likely to be salient enough to bias results. First, when streams are uncategorized, it is typically at the beginning of a streaming session before the streamer has loaded a game. Second, streamers have a strong incentive to categorize their streams accurately because it facilitates discovery. In addition, the *Twitch* rules of conduct prohibit intentional miscategorization of streams. Finally, streamers do not have an incentive to miscategorize their streams as this would be immediately observable by viewers upon entering the stream. While it is possible that some streamers unintentionally miscategorize their streams—for example if they switch games but forget to switch categories—such behavior is likely to be semi-random and infrequent relative to the number of correctly categorized streams.

Table 3-1. *Twitch* data summary statistics.

	Min	Max	Mean	S.D.
Category sessions				
Duration (hours)	0	869	0.66	1.68
Unique tag ids	0	43	1.46	1.21
Avg. concurrent viewers	0	1,605,231.7	10.41	313.05
Stream sessions				
Category sessions per stream	1	765	1.63	4.06
Duration (hours)	0	869	1.24	2.40
Unique games	0	66	1	0.53
Unique tags	0	43	1.56	1.33
Avg. concurrent viewers	0	981,598.1	10.26	300.24
Users				
Distinct days streamed on Twitch	1	504	14.65	34.04
Category sessions	1	46,013	33.68	217.65
Stream sessions	1	40,139	20.71	57.5
Unique games	0	1,261	3.23	6.17
Unique tags	0	223	2.29	3.68
Avg. concurrent viewers	0	243,503.0	7.77	291.30

CHAPTER 4

CONFUSION OR COHERENCE? HOW CATEGORY SPANNING BY LIVESTREAMERS REFLECTS COMPETING THEORIES OF VALUE

INTRODUCTION

An extensive body of work in organizational and economic sociology documents the tendency for audiences to penalize offerings that span multiple categories (Hsu and Hannan 2015; Negro, Koçak, and Hsu 2010; Zuckerman 2017). This penalty for categorically impure offerings has been found in contexts as diverse as wine ratings (Negro and Leung 2013), markets for financial capital (Zuckerman 1999; Leung and Sharkey 2014; Ruef and Patterson 2009), the labor market for Hollywood actors (Zuckerman, Kim, Ukanwa, and von Rittman 2003), and reviews of Hollywood movies (Hsu, Hannan, and Koçak 2009). For example, Negro and Leung (2013) found that wine producers who make both traditional and modern wines are devalued relative to specialist producers in either category, even though their wines fare equally well in blind taste tests.

A prevalent (and often implicit) explanation for why audiences penalize categorically impure offerings is rooted in a cognitive theory of categorization called prototype theory (e.g., Hannan, Pólos, and Carroll 2007; Hsu, Hannan, and Koçak 2009; Leung and Sharkey 2014; Negro and Leung 2013). The central claim of this explanation is that audiences are confused by offerings that span multiple categories because they defy simple classification and are thus more difficult to interpret. As a result of this confusion, categorically impure offerings are penalized relative to categorically pure offerings. For example, explanations rooted in prototype theory suggest that an evaluator will have more difficulty interpreting the identity of a wine producer who makes both traditional and modern wines ("what kind of vineyard is this?") and will penalize offerings from this vineyard even when their quality is equal to that of a specialist. Extending this logic, the penalty is more severe for combinations of categories that are atypical or distant as these are more confusing (Kovács and Hannan 2015).

Explanations rooted in prototype theory cast the penalty for category spanning as an "invariant behavioral tendency" (Zuckerman 2017), meaning that the penalty is a kind of cognitive bias inherent in the way we process categories that applies to all people and contexts. Casting the penalty for categorical impurity as an invariant behavioral tendency offers a compelling explanation for the seeming ubiquity of the penalty. And insofar as categorically impure objects are devalued *even after controlling for quality*, it seems reasonable to infer that this residual difference in valuation is due to cognitive bias.

However, as a general explanation of how people evaluate category-spanning offerings, prototype theory is problematic for two reasons. First, many empirical results from studies on categories and valuation are difficult to explain using prototype theory and even directly contradict the core claim that prototype theory makes about category-spanning offerings being penalized. Numerous empirical studies document situations where audiences not only vary in the extent to which they penalize categorical impurity (e.g., Kacperczyk and Younkin 2017), but also place a premium on category-spanning objects (Smith 2011; Pontikes 2012; Merluzzi and Phillips 2016; Sgourev and Althuzien 2014; Paolella and Durand 2016). Pontikes (2012), for example, found that venture capitalists recognize the enormous value to be gained by redefining market structure and thus place a *premium* on firms with ambiguous classifications. Such evidence directly contradicts the core premise of prototype theory, that (a) category spanning is penalized, and (b) it is invariant. As a cognitive explanation, prototype theory cannot be easily modified to accommodate systematic variation in valuations.[6]

[6] It is also worth noting that the posited confusion mechanism is also problematic. As Zuckerman (2017) points out, there is no general reason to believe that categorical combinations—even infrequent or novel ones—are confusing, or that confusion necessarily results in devaluation.

Second, prototype theory faces a number of theoretical challenges. In particular, it does not offer a compelling account of why certain categorical combinations become prototypical, nor does it offer an account of why prototypes might change over time. Prototype theory simply suggests that current prototypes reproduce themselves. While prototype effects may exist, prototype theory as a whole is not able to account for the range of categorical combinations observed in practice. These issues are particularly salient in settings where producers must adapt to rapidly changing environments. Prototype theory suggests producers should look to what has been typical in the past, rather than look to the future where opportunity lies. In this way, prototype theory offers an overly conservative account of strategic behavior and innovation.

While prototype theory overstates the case for people's aversion to category spanning, studies that document a premium for category spanning face similar limitations. One reason is that studies documenting a premium for category spanning often resort to preference-based explanations, i.e., that audiences (at least in some contexts), value category spanning for its own sake. This explanation is problematic because it does not provide a more general framework for adjudicating between when we should expect audiences to place a premium or penalty on category spanning. In much the same way that it is relatively straightforward to come up with examples where people prefer offerings that span multiple categories, it is relatively straightforward to come up with examples that combine categories in ways that are either incoherent or less useful than a more focused offering. A shortcoming of many empirical studies on category spanning is that such combinations are unlikely to be observed in practice, either because sophisticated producers readily intuit that some combinations are incoherent or value destroying or because these combinations are penalized so heavily that they quickly fail. Consider the aversion to startups that claim to be "Uber for *[service]*" or a "blockchain-enabled *[service].*" These combinations may have initially been met with some enthusiasm but are now posterchildren for ill-conceived ideas that misunderstand the underlying economics of the industry they are so keen to disrupt (Webb 2016; Higginson, Nadeau, and Rajgopal 2019). While some of these are good ideas and go on to be successful, the range of valued combinations is much narrower than suggested by a simple preference for generalists. A theory of categorization and valuation must be able to account for a generalist premium, but it must also place bounds on the range of combinations that are likely to be observed and valued.

This paper seeks to explain why some combinations of categories are more viable than others. To this end, I draw on the "theory-of-value" approach to categorization (Durand and Paolella 2013; Paolella and Durand 2016; Zuckerman 1999, 2017) to develop an alternative standard to typicality for constructing and evaluating category-spanning offerings. This standard, which I call *theoretical coherence,* focuses on the extent to which a combination coheres with an evaluator's theory of value. Importantly, what is considered theoretically coherent can differ from one person to another. This approach therefore provides a foundation for understanding heterogeneity in category schemes. In particular, we can see why producers facing the same broad market may choose to organize the category space in completely different ways. Compared to the standard of typicality, the standard of theoretical coherence provides a more robust foundation for understanding how categories constrain strategic choice and valuation. In particular, it also provides a more solid foundation for understanding why certain categorical combinations become prevalent and why these might change over time.

This analysis presented in this chapter overcomes two empirical challenges in studying theories of value. First, many settings are characterized by a dominant logic of categorical combination that precludes other potentially viable combinations. For example, stock market analysts hold strong beliefs about the value-destroying effects of diversification (Davis, Diekmann, and Tinsley 1994; Brealey and Meyers 2000; Zuckerman 2017). We are unlikely to observe multiple, viable theories of value operating in such settings and are thus unable to test whether there are viable alternative ways of structuring categories that could be valuable. What appears to be an aversion to atypical combinations could be explained more simply as the application of a particular theory of value. Using data from multiple settings introduces its own issues, as idiosyncratic factors like the range of offerings being considered introduce confounding effects.

Using a novel dataset from the videogame livestreaming platform *Twitch*, I am able to overcome this first limitation. A unique feature of this setting is that there are multiple viable theories of value being used to organize the category space (in this case, videogames) and I am able to observe these theories of value. *Twitch* is a popular video live streaming platform focused on video games, where "streamers" broadcast live video of themselves playing a game to an audience of viewers. *Twitch* consists of many niche communities that play similar games but in very different ways. Producers on *Twitch* thus have a range of viable strategies at their disposal to appeal to different audiences. By leveraging the range of producer strategies on *Twitc*h, I show how different theories of value lead to different ways of organizing the category space (games in a streamer's repertoire). In this paper, I focus on two such play styles, speedrunning and eSports, because they both communities share an emphasis on extreme displays of skill and expert levels of competitive play. Yet, seemingly subtle differences in the playstyles of these mean that their audiences differ in what they consider a coherent combination.

A second empirical limitation in testing theoretical coherence is that the theory of value cannot be measured using observed combinations of categories. Observed combinations can only tell us what is typical. Yet an atypical combination can be coherent and a typical combination can be incoherent. Measuring theories of value thus requires a separate way of measuring the conceptual relationships among categories. To overcome this second limitation, I use data from a separate platform, *Reddit*, to identify the theories of value espoused by these communities and generate measures of theoretical coherence to place bounds on the range of combinations that will be valued by audiences.

In the following section, I provide an overview of prototype theory and theories of generalist premiums and explicate their limitations in understanding which combinations of categories are viable. I then introduce the theory of value approach, and specifically its implied standard of theoretical coherence, as a way to address these shortcomings. The following section introduces *Twitch*, the main setting for the study, and details the empirical strategy. I then present results to validate key implications of my theory.

THEORY

Prototype theory

Prototype theory, as invoked by economic and organizational sociologists, makes two main claims about category spanning objects: (1) that they are confusing and (2) that they are devalued because of this confusion (Hsu, Hannan, Koçak 2009; Hannan, Pólos, Carroll Kovács and Hannan 2015; Leung and Sharkey 2014; Negro and Leung 2013). This view holds that categories are central to people's expectations of offerings and how they make sense of them. Because categorically impure offerings defy simple classification, audiences find them challenging to interpret and they consequently devalue them. Producers and offerings that span multiple categories are thus seen as less appealing than more focused producers and offerings.

This explanation is rooted in a cognitive theory of how people represent and process categories. Given its cognitive roots, the theory understands the penalty for categorical impurity as an invariant behavioral tendency that operates across audiences and contexts (Zuckerman 2017). Prototype theory sees categories as represented by a prototype,[2] an abstract or tangible ideal offering that is the best exemplar of the category. Category membership is evaluated based on similarity (in some feature space) to this prototype; objects that are more similar to the prototype are considered to be more representative or prototypical of the category. In this way, category membership is "graded" and categories are conceptualized as fuzzy sets (Hannan 2010). For example, people report robins to be more representative of the category *bird* than turkeys and penguins (Rosch 1975). Robins therefore have a higher grade of membership in the bird category.

Objects that span categories possess some but not all features of the prototypes representing the spanned categories. As a result, they are less prototypical of the respective categories and their grade of membership in those categories is reduced. Being distant from prototypes, categorically impure objects are more difficult to classify than pure ones and are in turn more challenging to make sense of. The extent to which category spanning objects are confusing increases with the number of categories spanned and when the categories spanned are more distant in feature space (Kovács and Hannan 2015).

Limitations of prototype theory to be addressed

Although prototype theory provides a compelling explanation for why the multicategory discount seems so prevalent, casting the multicategory discount as an invariant behavioral tendency is problematic for three reasons: (1) if the discount is invariant, it cannot account for observed variation in demand for categorical purity, especially premiums for category spanning;

and (2) there are reasons to doubt the theory's strong causal claims about confusion and devaluation (for more discussion of these and other issues with prototype theory, see Zuckerman 2017); and (3) it does not provide a general account of which combinations will emerge as prototypical and why these combinations change over time. A broader theory is needed that can account for both the seeming ubiquity of the multicategory discount but also account for variation in demand for categorical purity while also providing insight into the emergence and change of categorical combinations observed in practice.

Variation in tolerance of categorical impurity. If the penalty for categorical impurity is driven by an invariant behavioral tendency, then why do we observe such variety in demand for purity (Zuckerman 2017)? On the one hand, some studies find a generalist premium (e.g., Sgourev and Althuzien 2014; Merluzzi and Phillips 2016; Smith 2011). For example, Merluzzi and Phillips (2016) find that elite MBA graduates seeking jobs at investment banks are more likely to receive an offer and are given higher starting bonuses when they have more general experience compared to peers with profiles specialized in investment banking. On the other, studies identify several sources of variation in demand for purity. In a study of software organizations, Pontikes (2012) finds that venture capitalists place a premium on organizations with ambiguous classifications, whereas consumers penalize such organizations. In a study of the early stages of the credit rating agency R. G. Dun and Company, Ruef and Patterson (2009) find that credit reporters initially imposed few penalties on category-spanning businesses, but that these penalties became more pronounced as the credit reporting system became institutionalized. Other studies find that audiences are relatively unconcerned with categorical impurity when the producer has already demonstrated capability and commitment (for a more general discussion, see Zuckerman 2017; Zuckerman et al. 2003; Kacperczyk and Younkin 2017; Sgourev and

Althuizen 2014). These findings directly contradict the main prediction of prototype theory—that category spanning objects will be devalued.

Confusion. Are category-spanning objects really so difficult to understand? Original formulations of prototype theory by cognitive scientists (Rosch and Mervis 1975; Rosch 1978) have little to say about categorical combinations and do not make any claims about how they relate to understanding or valuation. While this absence does not itself invalidate the claim that category spanning objects cause confusion, it is a strong claim that lacks theoretical warrant and, to my knowledge, has not been directly tested. There are nevertheless good reasons to doubt that category spanning necessarily entails confusion and confusion necessarily entails devaluation. In particular, category spanning objects can pique our curiosity and "intrigue us to take a closer look" (Zuckerman 2017: 40) and can be interpreted as a sign of creativity (Sgourev and Althuizen 2014) or innovation (Pontikes 2012). Even the link between confusion and devaluation may not be as straightforward as suggested by the versions of prototype theory invoked by economic sociologists and organizational theorists. Pocheptsova, Labroo, and Dhar (2010) report that over 200 studies find metacognitive difficulty reduces valuations of an object, but also find that the effect of metacognitive difficulty is contingent on the consumption domain, which they summarize as follows: "In the domain of everyday goods, metacognitive difficulty reduces the attractiveness of a product by making it appear unfamiliar. However, in the context of special-occasion products, for which consumers value exclusivity, metacognitive difficulty increases the attractiveness of a product by making it appear unique or uncommon" (Pocheptsova, Labroo, and Dhar 2010: 1059). By explicating the causal chain of reasoning that relates confusion to valuation, we can now see more clearly how prototype theory's claims about confusion and value necessarily entail the application of a theory of value.

As a preliminary step to developing a broader theory of value that can subsume prototype theory, it is useful to make explicit some problematic assumptions on which prototype theory relies on to make its claims about confusion. Recall that prototype theory (as invoked to explain the penalty for categorical impurity) posits that value is a direct function of an object's grade of membership (in a positively valued category). Category-spanning objects are conceptualized as sitting between prototypes in feature-space and therefore necessarily have a lower grade of membership in the constituent categories than fully-fledged members. Such objects are believed to be less able to meet expectations associated with the constituent categories.

But what are the salient features of prototype that affect valuation and on what basis is distance evaluated? Absent a theory that points to salient features and how to compare them, this "distance" argument runs into difficulties. Consider the following counterexample. Guppies are neither prototypical of the category "fish" nor the category "pet," but are highly prototypical of the composite category "pet fish" (Osherson and Smith 1981). The way prototype theory is invoked to explain valuation would suggest that guppies are prized neither as pets nor as fish, but are highly prized as pet fish. But this is awkward for the main thrust of the theory: that category spanning objects are devalued.

Prototype theory's claims about distance between prototypes, confusion, and their relation to value also rely on unspecified theories of category representation and processing that are not warranted by the original formulations of prototype theory. Rosch explicitly disavows both interpretations of prototype theory:

> … prototypes themselves do not constitute any particular model of processes, representations, or learning. This point is so often misunderstood that it requires discussion… To speak of a prototype at all is simply a convenient grammatical

> fiction; what is really referred to are judgments of degree of prototypicality… For natural-language categories, to speak of a single entity that is the prototype is either a gross misunderstanding of the empirical data or a covert theory of mental representation. (Rosch 1978: 16)

Making these theories explicit helps us understand why casting the multicategory discount as an invariant behavioral tendency is problematic and why a more flexible theory of categorization is required. Of course, lack of theoretical warrant does not in itself render a claim false. It does, however, suggest that (i) there is a more restricted role for prototype theory and confusion in categories, and that (ii) the implicit theories of category representation and processing undergirding the confusion and devaluation mechanisms be made explicit and subject to empirical testing.

Theory-based theories of categorization and the theory of value approach

To provide intuition for the theory-of-value approach, consider the following example:

> A somewhat unusual, but nonetheless useful, example arises from an old puzzle of biblical scholarship, the dietary rules associated with the abominations of Leviticus, which produce the categories clean animals and unclean animals. Why should camels, ostriches, crocodiles, mice, sharks, and eels be declared unclean, whereas gazelles, frogs, most fish, grasshoppers, and some locusts be clean? What could chameleons, moles, and crocodiles have in common that they should be listed together? That is, what is there about clean and unclean animals that makes these categories sensible or coherent? (Murphy and Medin 1985: 289)

Mary Douglas famously illustrates how the book of Leviticus organizes dietary laws based on an implicit theory of purity, one that reflects the concerns of a pastoral community. Someone who wishes to be pure and who subscribes to this theory of purity can be expected to organize and consume their foods according to this theory. But this is just one way of organizing foods. Others may organize foods according to the principles laid out by Jainism, freeganism, or sodium content. In some cases, people may share the same goal (purity, health) but espouse very different theories about how foods relate to those goals, and organize foods in correspondingly different ways. In other cases, people may have different goals but organize the category space in the same way. Understanding this theory of value is essential not only to the consumer, but also to a producer who wishes to sell foods to such a consumer so that they may combine foods in a value-creating way.

This example captures the essence of the theory of value approach:[7] the premise that people have objectives and theories about how to best achieve those objectives (Durand and Paolella 2013; Paolella and Durand 2016; Zuckerman 1999, 2004, 2017; Zuckerman et al. 2003). In contrast to prototype theory, the theory of value approach sees categorization as a flexible tool that organizes offerings in a way that is conducive to achieving particular goals. This is evident in the way that both producers and audiences use categories strategically. For example, producers strategically claim membership in beneficial categories and try to shape categorical boundaries to meet their goals (Granqvist, Grodal, and Woolley 2013; Hsu and Grodal 2015).

[7] Durand and Paolella (2013) delineate three models of categorization developed by cognitive scientists: prototype theory, the causal-model approach and the goal-based approach. Whereas the causal-model approach to categorization emphasizes prior knowledge about processes, the goal-based approach emphasizes an actor's goals. The theory-of-value approach can be understood to encompass both in that it considers both an actor's goals and their theories, broadly understood, about how to achieve those goals.

After all, if the construction and valuation of categories were a purely behavioral phenomenon, there would be little room for strategy.

The primary implication of the theory-of-value approach for understanding the role of categories in valuation is that a preference for categorical purity or impurity is driven by a particular logic. Evaluators may favor offerings that span multiple categories over pure offers, when such category spanning aligns with their theory of value. In the market for corporate legal services, for example, Durand and Paolella (2013) find that clients value category-spanning law firms more highly when their needs are complex because they are thought to be more capable of the greater breadth of issues that arise in a complex case. Contrary to theories that focus on generalist premiums, these law firms do not seek out generalists simply because they value category spanning in and of itself, but because their theory of value recognizes that participation in relevant categories is more conducive to achieving their complex goals.

However, an important implication implicit in prior work on the theory of value approach is that only *particular* combinations of categories will be valued. Theories of value have an internal logic about which combinations of categories work well together. Some combinations will cohere with the theory of value, while other combinations will be incoherent from the perspective of the theory. For example, an international shipping company would likely value the services of a law firm spanning the areas of maritime law, international law, and corporate law. By contrast, it is less likely to value the services of a law firm spanning maritime law, entertainment law, and civil rights law. Thus, theories of value not only shape how people categorize offerings, but also how they structure the broader space of categories and perceive relationships among categories.

By explicating a particular theory of value, we can understand which combinations of categories will be coherent with that theory, and thus which specific combinations of categories will be valued. Importantly, theoretical coherence is orthogonal to typicality: typical combinations can be incoherent and atypical combinations can be coherent. In nascent or turbulent settings where prototypes may not exist, this logic can be extended to encompass new and unseen offerings. Relative to theories that posit people either value or devalue categorically impurity for its own sake, the theory of value approach provides more precise bounds on the range of categorical combinations that will be valued or devalued.

The theory of value approach enriches prototype theory by addressing a key conceptual issue mentioned earlier: theories of value identify salient dimensions on which offerings are compared and provide a basis for reasoning about similarity. Recall that prototype theory asserts that objects that are more "distant" from a prototype (in some abstract feature space) are more challenging to interpret and thus more likely to be devalued. But how do people choose what features to compare and how do they evaluate "distance"? As an example, consider how a person might evaluate the substitutability of two items. Is vanilla or coffee a better substitute for chocolate? Either might be acceptable in an ice cream, but coffee would be a far better substitute in a chili dish. These processes can be better understood in terms of the (often implicit) theories that people invoke to draw attention to salient features of an offering (e.g., color, taste, texture, price) and how they should be evaluated.

The standard of theoretical coherence implies two general propositions about how coherence relates to value. A minimal criterion for valuing categorical combinations is that they should be coherent according to the espoused theory of value. In other words, a combination that appears incoherent from the perspective of established categories can be valued if it is theoretically coherent. In cases where audiences are generally inclined to value categorical combinations, they are expected to value more highly those that are most coherent from the perspective of their theory of value, and in cases where audiences are generally inclined to devalue categorical combinations, they are expected to be more tolerant of combinations that are more coherent. Hence, the following general propositions:

> ***Proposition 1a:*** *Audience members are more likely to value categorical combinations that are coherent from the perspective of their theory of value, and less likely to value categorical combinations that are incoherent.*

> ***Proposition 1b:*** *Audience members are more likely to penalize categorical combinations that are incoherent from the perspective of their theory of value and less likely to penalize categorical combinations that are coherent.*

The final shortcoming of prototype theory that must be addressed is the issue of change. How does theoretical coherence explain the emergence and change of typical combinations? I return to this question in the discussion section, where I reflect on the relationship in light of my results. I will then build on this perspective in the next chapter.

EMPIRICAL SETTING AND APPROACH

I now turn to examining the construction and valuation of *Twitch* streamers' game repertoires, and thereby illuminate how theories of value constrain strategic choice and valuation as specified in propositions 1a and 1b.

An important feature of *Twitch* that makes it a particularly useful setting for studying theories of value is that it is a setting where multiple theories of value are present. The presence of multiple theories of value implies that there are multiple viable ways for producers to combine categories. Prior research on categorization and valuation has typically been conducted in settings with a single, dominant theory of value. One manifestation of such studies are settings where there is a single evaluator. The taken-for-grantedness of hegemonic theories of value obscures the alternative constraints on categorical combinations that would occur under counterfactual theories of value. By comparing competing theories of value within a single platform, my *Twitch* data afford an unusually close comparison of competing theories of value, which allows me to control for many confounding factors. In particular, by narrowing my focus to two types of competitive videogaming, I am able to study these differences among producers and audiences operating in the same market context (*Twitch*), who are combining the same set of offerings (videogames), and who share the same broad goals (elite displays of skill).

Data

In addition to the quantitative data used in this study, my knowledge of *Twitch*, eSports, speedrunning, as well as the videogaming and livestreaming industries more generally, is informed by several types of qualitative data, including community discussions, interviews, and informal participant observation. My observations include watching more than one hundred hours of streams on *Twitch* as well as other videogame livestreaming platforms. These streams included eSports and speedrunning streams, as well as a variety of other types of streams. I read hundreds of forum posts in communities dedicated to eSports, speedrunning, videogaming, and livestreaming. To understand the strategic considerations driving the composition of game repertoires, I read numerous guides providing advice to streamers on how to grow their channels, including the official best practices provided by the *Twitch Creator Camp*. I also attended community events. At TwitchCon 2018 (the annual *Twitch* convention held in San Jose, California), I spoke to streamers, viewers, developers, and eSports competitors and industry professionals. I also spoke with several people from adjacent industries, such as US Army Rangers seeking to recruit eSports competitors and private banking representatives catering to professional livestreamers and eSports competitors. I also attended PAX East 2019, another major gaming convention. Although I do not conduct a formal analysis of these qualitative data, I use them throughout my analyses to guide interpretation of my data and validate my empirical framework.

Game repertoires and category spanning on *Twitch*

The empirical approach taken in this chapter rests on the premise that competing theories of value will be reflected in the construction and valuation of streamers' game repertoires. By a streamer's repertoire, I am referring to the set of games that a streamer plays for their audience. Some streamers play only a single game and thus have a "pure" repertoire. As Figure 4-1 shows, however, complete purity is relatively rare on *Twitch*. Most streamers have multiple games in their repertoire, with a substantial number of streamers having more than 50 games. Thus, at any one time, a streamer may be associated with multiple games.

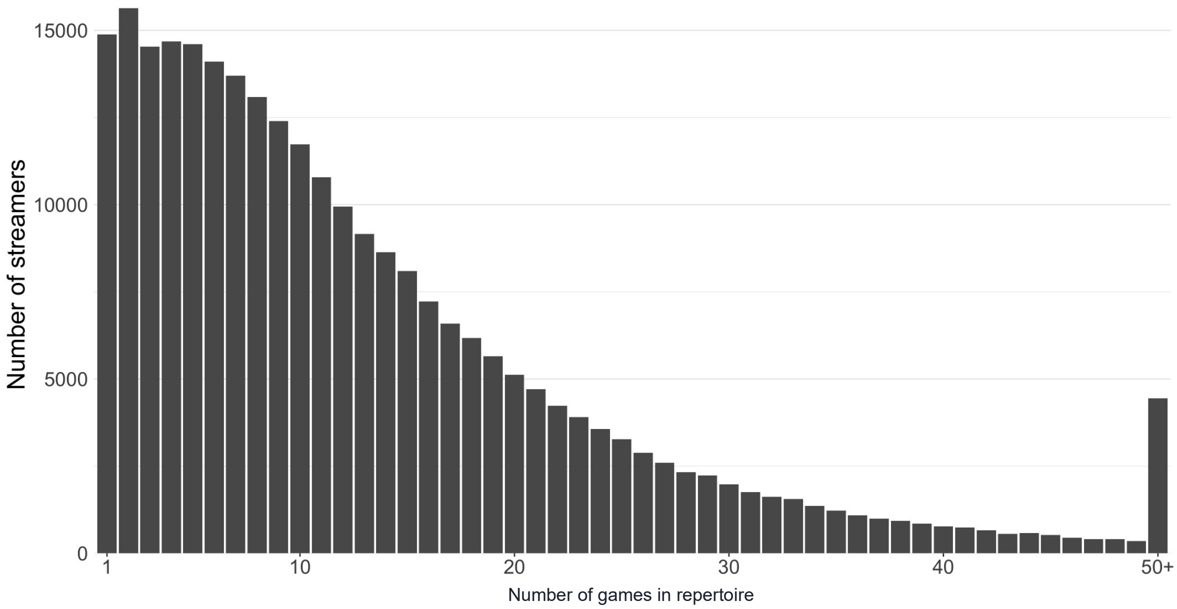

Figure 4-1. Distribution of repertoire size (all streamers). Repertoire size is the number of unique categories a streamer has streamed in during the observation period. This distribution is for all streamers.

Games constitute one of the most salient categories on *Twitch,* both in terms of how the platform is structured as well as how users understand and interpret streamers' offerings. That games are central to the *Twitch* experience is evident in the fact that one of the few ways the platform organizes streams is by game. As discussed in Chapter 2, each game is given its own category in the *Twitch* directory.

To shed light on the original question of why some combinations of categories are more viable than others, I examine how audiences value particular combinations of games in streamers' repertoires. At first glance, it may seem odd to equate games with categories, as categories are typically thought of as more akin to genres, whereas a videogame is typically considered a product that belongs to a genre. Games can and are certainly grouped into genres such as shooting games, strategy games, and racing games.

To clarify how repertoires comprised of various games relate to category spanning, it is useful to see the games in a streamer's repertoire not as products that the streamer is reselling but as a *context* in which the streamer delivers value via a service. Although many streamers play the same game, their offerings are not commodities, but are provided in highly differentiated ways, along such dimensions as personality and playstyle. For example, for two streamers who each play the game *Fortnite,* one may attract viewers because they demonstrate high levels of skill in the game, whereas the other might attract viewers because of their personality. Thus, the value that a viewer gets from watching a streamer play a certain game cannot be reduced to the game's genre, but must be understood in terms of the service the streamer is delivering via the game.[8] As such, whether a repertoire combines games in a coherent way cannot be evaluated without knowing whether the streamer is able to competently deliver their service through those games.

Competing theories of value on *Twitch*

I now describe competing theories of value governing the construction and valuation of game repertoires on *Twitch.* I begin by examining a common distinction made between "single-game" and "variety" streamers. This discussion helps to contextualize streamer strategy and audience valuation on *Twitch.* I then explain how the constraints on strategic choice and valuation facing these streamers can be better understood in terms of skill and relational theories of value. Given that audiences and producers operating under skill and relational theories of value operate under broadly different market contexts and have different goals, it is not surprising that they would combine games differently. To address this issue, I then narrow my focus to two specific types of

[8] Viewers may of course only be interested in particular game genres, such as shooting games, and only watch streamers who play such games. In such cases, genre would be a minimum but not sufficient condition for selecting a streamer to watch (Zuckerman 2017). However, this does not invalidate the central point, which is that the categorization and valuation of streamers cannot be reduced to a feature of the game, like its genre, but is based upon a theory about the value of the streamer's service.

skill-based streamers: eSports competitors and speedrunners. This allows me an unusually close comparison of producers and audiences who operate in the same market context (competitive videogaming on *Twitch*), combine the same types of offerings (games), and who share similar goals (demonstrations of elite skill). By exploiting the differences in the theories of skill governing these communities, I am able to show how nuanced differences in theories of skill imply consequential differences in which combinations of games will be viable.

Single-game vs. variety streamers

A common distinction made by streamers and viewers on *Twitch* is between *single-game streamers* and *variety streamers*. As the names imply, single-game streamers primarily play a single game[9], whereas variety streamers play a variety of games. The decision to be a single-game or variety streamer is of strategic importance to streamers because the first transition from playing a single game to adding another game to one's repertoire is widely thought to have significant performance implications. For example, the negative consequences of deviating from purity are echoed by one streamer, who describes how he lost half of his subscribers when he started playing another game:

> I had these 38,000 followers, all of them have only watched me play *Mario Maker* for my entire career. And then when I basically cut my stream in half, so that I played *Mario Maker* in the first half and [a shooting game in the second half]. Yeah, basically I saw a lot of people that would watch me only for the *Mario Maker* in that time frame. They would basically unsubscribe.

[9] Streamers who specialize in a single game franchise might also be considered single-game streamers. A game franchise refers to games in the same series. For example, *Super Mario World* and *Super Mario 64* are both part of the *Super Mario* franchise.

That transitioning to variety streaming generally incurs a penalty is a widely held belief among *Twitch* users, as illustrated by the following quote from *Reddit* user sadpandadag (2015) describing why purity is beneficial:

> [single-game] streams have an extreme advantage… Individual games attract audiences. There are people who care about watching specific games or popular new releases. Tapping into the community of viewers around these games gives them access to a large potential viewership that is largely guaranteed as long as they stick to those games. [Streamers who play multiple games] do not have that benefit. They may temporarily attract viewers by playing one of those games, but are largely unable to retain that increased viewership when they ultimately move on to a different game.

These quotations illustrate a lay theory of value corresponding to a multi-category discount, with an underlying logic that is not incompatible with prototype theory.

Despite these challenges, streamers also recognize that making the transition to variety streaming can be beneficial. In a video giving advice to new streamers, Retro Ali, a streamer of Japanese role-playing games and retro games with over 34,000 subscribers on *Twitch*, explains how broad repertoires are conducive to building audiences: "When you're trying new games you can reach a whole new audience that you never reached before" (Retro Ali 2017). Concurring with this view, sadpandadag (2015) adds that broad repertoires also minimize risk from fads and fashion cycles[10]: "[when playing multiple games] you'll build an audience of people who

[10] This logic is familiar from organizational ecology, where high levels of variability in coarse-grained environments favors generalists over specialists (Hannan and Freeman 1977; Freeman and Hannan 1983; see also Reagans and Zuckerman 2008).

genuinely enjoy watching you play games and this severely mitigates the danger of losing your viewership based on the games you choose to play."

On the surface, the distinction between single-game and variety streamers appears to map onto a distinction between categorical purity and impurity—or a preference for specialists vs. generalists. However, the quotes also begin to suggest why casting the valuation of category-spanning producers in terms of penalties/premiums for specialists and generalists is problematic. In particular, the variation in tolerance for variety streaming does not appear to be related to confusion or atypicality and therefore cannot easily be explained by prototype theory. The quote by sadpandadag suggests that variety streamers' audiences are (at least in some instances) drawn to variety streamers because of the streamer's personality or the audience's relationship to the streamer.

Skill vs. relational theories of value

Although the distinction between single-game and variety streamers is a useful starting point for gaining a deeper understanding of the constraints that streamers face, further investigation reveals that the single-game/variety distinction (and by extension the specialist/generalist distinction) is too blunt. In particular, the distinction does not provide an explanation of why some combinations of games are more viable than others. This sentiment is echoed by SimCopter1 (2018), who says: "I don't call myself a variety broadcaster because it doesn't really describe anything. It does not help you stand out." Variety-streaming, like theories of generalist premiums, offers little guidance in the construction and valuation of repertoires.

To better understand the constraints imposed on single-game and variety streamers, it is useful to first understand two general dimensions on which viewers evaluate a streamer: the streamer’s skill versus the streamer’s personality and the relational aspects of livestreaming. As another streamer explains:

> There's really like two types of streamers. You got your streamers who are really good at video games, then you got your streamers who are like personality driven, who suck at video games. We fall into the latter. We suck. We're not good at games at all. We don't pretend to be good at games. We're just all about having a great time.

These two types of streamers may be thought of as governed by two distinct theories of value, one emphasizing skill and the other emphasizing personality or relationships. Rooting the evaluation of streamers in skill- and relationship-based theories of value provides a broader foundation for understanding not only preferences for single-game versus variety streamers, but also the kinds of combinations that audiences are willing to accept.

Audiences who apply a skill-based theory of value can be expected to display greater preferences for purity compared to audiences who apply a relationship-based theory of value. To see why this is the case, consider how demand for high levels of skill creates pressure for purity whereas demand for relationships creates pressure for impurity. Insofar as displaying high levels of skill—both relative to oneself and others—requires specialization, streamers are less likely to be able to display high levels of skill in multiple contexts. By contrast, consider the importance of interacting in multiple contexts for deepening relationships. While two people who only interact in the context of work may grow closer, their relationship is more likely to deepen when they interact in a variety of contexts, such as meeting each other’s families or engaging in mutual hobbies. Audiences who apply a relationship-based theory of value are more likely to place a premium on variety streaming as relationships are strengthened through interaction in multiple contexts.

To be sure, the distinction between skill- and relationship-based streamers is a crude one. Most viewers demand that streamers display some minimum level of both skill and personality. Nevertheless, it is a useful distinction because it forces a shift from focusing on categorical purity vs. impurity (or specialists vs. generalists) to seeing particular logics of combination that guide the construction and valuation of repertoires. In particular, audiences who apply a skill-based theory of value will consider game repertoires more coherent when there is a high degree of skill overlap in the games. Although audiences who apply a relationship-based theory of value are more diffuse in the combinations of games they are willing to accept, they do not give streamers a *carte blanche*. Some games provide more suitable contexts in which to develop the streamer-audience relationship. For example, it is easier for the streamer to interact with their audience while playing a casual game compared to a highly competitive game where more focus is required. To better examine how theories of value shape the types of combinations audiences are willing to accept, I now turn to a narrower comparison between two skill-based theories of value.

Competing theories of skill: eSports competitors and speedrunners

The distinction between skill- and relationship-based streamers provides a stark contrast in logics for constructing and evaluating repertoires. To provide a more direct comparison and demonstrate how even nuanced differences in theories of value affect the way producers and audiences organize category space, I now narrow my attention to two types of skill-based streamers on *Twitch*: eSports competitors and speedrunners. Both eSports and speedrunning are forms of competitive videogaming that involve demonstrations of elite skill and expertise in a game, not unlike traditional sports. However, eSports and speedrunning differ in their goals and approaches to gameplay. As a result, they are characterized by different theories of skill.

While eSports streamers and speedrunners are by no means representative of the larger population of streamers, let alone other kinds of producers, narrowing my analysis to these two groups affords me a unique opportunity to test the theory of value approach by eliminating many confounding factors that could explain differences in valuation. In particular, it is not surprising that we would find differences in valuations if we were to compare audiences and producers in different market contexts or with different goals. Comparisons across market contexts are particularly unconvincing because producers and audiences face entirely different offerings. For example, the logic of combination in financial markets cannot readily be applied to the market for MBA candidates. By focusing on eSports and speedrun streamers, I am able to compare producers operating in the same market context (*Twitch*), with the same offerings at their disposal (games), and who share similar goals (displaying expertise in competitive videogaming). This strategically selected sample thus allows me to make an unusually close comparison of nuanced differences in theories of value.

In eSports, individuals or teams compete against each other in organized tournaments. Once a niche activity limited to only the most diehard gamers, eSports has rapidly grown in popularity over the last decade and become a mainstream activity and is now similar to traditional sports in terms of its organization and professionalization. In the same way traditional sport competitions revolve around individual sports like football and basketball, eSports competition revolves around specific games, like *Fortnite* and *Rocket League.* Major game franchises often feature both amateur and professional leagues. On the amateur side, an increasing number of high schools and colleges have created eSports teams and even offer scholarships to attract players.[11] On the professional side, eSports players sign contracts with teams to train and compete full time. Professional eSports tournaments often feature significant cash prizes. For example, in 2019, 18 teams competed in The International, an annual competition for the game *Dota 2*, for a prize pool of $34 million USD.

Speedrunning is also a form of competitive videogaming, but one that takes a very different approach to games from eSports. Whereas eSports competitors play against others and seek to defeat their opponents, speedrunners attempt to complete a game (or a part thereof) in the fastest time possible. For example, someone who speedruns *Super Mario 64* may seek to set a record for beating the full game, or they may seek to set a record for a particular level. The website *speedrun.com* serves as the de facto leaderboard for the speedrunning community. Players who wish to record their times must record a video of their attempt and submit it to *speedrun.com*, where the community verifies the integrity of the attempt. Setting new records requires impeccable timing and pixel-perfect precision. A single attempt can range from minutes to tens of hours or more to complete. A single button mis-press or timing error that results in fractions of a second wasted is likely to result in a failed attempt at a record, after which they must start from the beginning. Speedrunning is a unique form of gaming in that it often uses unconventional methods to minimize time. In particular, speedrunning often involves the use of glitches or exploits that allow players to warp through the game and skip entire sections. Such tactics might be considered cheating in other contexts, but are highly esteemed as novel innovations in speedrunning.[12] Another type of speedrunning is called a tool-assisted speedrun,

[11] As of 2022, the National Association of Collegiate Esports lists 195 colleges and universities in the United States with varsity eSports teams. A number of schools, such as University of California Irvine and Ohio University provide scholarships to eSports competitors.
[12] Speedrun attempts can be performed under many different types of constraints. For example, *glitchless* speedruns do not allow the use of glitches or exploits. Games thus typically have several different categories for which times can be submitted.

where the player uses tools to slow down the game and achieve a theoretically optimal run. In this way, speedrunners are communities of user-innovators (von Hippel 2007) with a culture reminiscent of early hacking communities.

Twitch is a cornerstone of both the eSports and speedrunning communities. *Twitch* and its predecessor, *Justin.tv*, were instrumental in popularizing these activities by making broadcasts publicly available. Individual eSports competitors and speedrunners often stream on *Twitch.* While tournaments and competitions are featured on *Twitch* by the organizers*,* the focus here is on individual streamers.

Given that both eSports and speedrunning require mastery of videogames, they share many common skills. For example, both eSports competitors and speedrunners must have excellent mechanical control and intimate knowledge of the games they play. However, because of their different goals and methods for achieving them, the skills required to be competitive also differ. For example, because eSports competitors often compete in teams, they typically require excellent communication skills, whereas speedrunning is nearly always a single-player endeavor. By contrast, because speedrunners often seek new glitches and exploits, certain programming skills can be valuable, whereas such skills are not directly relevant to eSports. In subsequent sections, I explore these differences in greater detail.

Insofar as audiences on *Twitch* are drawn to competitive videogaming like eSports and speedrunning because of the high levels of skill on display, all other things equal, it is reasonable to assume that they will penalize streamers who display lower levels of skill. In the first instance, one might then expect the construction and valuation of such streamers' game repertoires to be driven by demand for purity, meaning that audiences place a premium on specialists and penalize streamers who play multiple games. Regardless of their demand for purity, audiences (and by extension streamers) can nevertheless be expected to apply a theory of skill that associates skill in one game to skill—or lack thereof—in other games (such theories of skill are at the heart of the typecasting challenge in Zuckerman et al. 2003). Consequently, if eSports and speedrun streamers differ in the theories of skill being applied to them, this would imply different ways of organizing the game space. This difference manifests because the basis for skill in these two domains is different, and in turn implies different logics of categorical combination that should be reflected in the games that streamers add to their repertoires and the consequences of those choices for their viewership numbers.

These differences in theories of value imply that speedrun and eSports streamers will evaluate the theoretical coherence of a given repertoire differently. For example, *Doom Eternal* and *Fortnite* are two games that are often combined by streamers on *Twitch*. Someone who applies the eSports theory of skill might consider these games a coherent combination as both are amenable to eSports competition and have similar mechanics. By contrast, someone applying the speedrunning theory of skill might consider such a combination to be incoherent because *Doom* can be speedrun, whereas *Fortnite* cannot because it is a multiplayer game. This also helps us understand why a combination like *Doom Eternal* and *Super Mario*—which is atypical and to an outsider would seem to be a case of unrelated diversification—can actually be considered theoretically coherent by someone who applies the speedrunning theory of skill. Both *Doom* and *Super Mario* can be speedrun, and are in many ways similar from the perspective of a speedrunner. For example, general speedrunning skills like glitch hunting can be applied to both games. However, the kinds of mechanics required to be competitive in a *Doom Eternal* eSports tournament have little carry over to *Super Mario*. The empirical approach taken in this paper leverages the presence of eSports and speedrunning theories of skill on *Twitch* to demonstrate how these theories of value affect the perceived coherence and valuation of streamers' game repertoires.

Empirical approach

My empirical approach is designed to systematically identify and test significance of the distinction between the two skill-based theories of value described in the previous section. The approach is divided into two steps. I begin by developing and validating a general empirical framework for identifying theories of value and measuring coherence of categorical combinations. This approach uses word embedding models trained on community discourse to uncover latent theories of value. Applying this approach to text data from two prominent eSports and speedrunning communities, I identify distinctive elements comprising their theories of skill. The results of the word embedding models are then used in my measure of theoretical coherence. In the present context, I create separate measures of coherence with regard to the eSports and speedrunning theories of skill.

In the second step, I test the relationship between theoretical coherence and audience valuations using panel data on eSports and speedrun streamers, as entailed by propositions 1a and 1b. For these analyses, I analyze the entire population of streams by eSports and speedrun streamers on *Twitch* over a period of more than two years. In these analyses, I also explore the relationship between coherence and typicality.

EMPIRICAL FRAMEWORK FOR IDENTIFYING THEORIES OF VALUE AND MEASURING COHERENCE

In this section, I develop and validate an empirical framework for identifying theories of value and measuring the extent to which categorical combinations cohere with a particular theory of value. A key advantage of this approach is that it does not rely on the observed co-occurrence of categories. The coherence measure is thus conceptually and empirically distinct from typicality, which in turn allows me to test whether coherence has an effect even after controlling for the typicality of combinations. Although I develop the framework in the context of eSports and speedrun streamers, the framework is readily generalizable to any setting where the researcher has access to text data reflecting the theories of value of the audience of interest.

Using community discourse to identify latent theories of value

Communities operate on the basis of shared understandings (Wenger 1998; Wohl 2015). Insofar as members of a community seek to coordinate around any kind of valuation or exchange, their shared understandings must also encompass theories of value. For example, in the case of competitive eSports and speedrunning, the communities around these activities must develop shared metrics for establishing and legitimating a ranking system that recognizes some competitors as better than others. Without shared metrics, the different performances of competitors would be incommensurable (Espeland and Stevens 1998). That community members share theories of value does not preclude heterogeneity or even disagreement in individual theories of value. As Goldberg (2011: 1397) succinctly summarizes the difference, "Sharing an understanding does not necessarily imply having the same opinions but rather agreeing on the structures of relevance and opposition that make symbols and actions meaningful."

A natural entry point for understanding a community's theory of value is to examine its discourse. Discourse proceeds on the basis of shared understandings, and salient theories of value should be reflect in written and oral discussions. However, little—if any—discourse can be expected to explicitly state theories of value. Even in situations where theories of value are stated or discussed explicitly, no single instance of such a discussion nor a single individual can claim to comprehensively codify the theory of value a community. Each document offers only a potential trace of insight into the deeper logic guiding valuation. To identify latent theories of value in a text thus requires a systematic approach that can uncover high-level structure among latent concepts.

Method

One such method is word embeddings, a method commonly used in natural language processing that has been shown to accurately encode relationships among latent concepts. In the social sciences, they have recently been used to study concepts, including political ideology (Rheault and Cochrane 2020), culture (Kozlowski, Taddy, and Evans 2019), and the intersectionality of race and gender in abolitionist discourse (Nelson 2021).

Word embedding models represent words as vectors in a high-dimensional space (Mikolov et al. 2013). Word embeddings reflect the Distributional Hypothesis, famously summarized as "you shall know a word by the company it keeps" (Firth 1957). The vectors are arranged such that words that appear in similar contexts are closer in the embedding space. As a result, word embedding models have the notable property that the geometry of the embedding space accurately encodes complex semantic relationships among words.

A classic example demonstrating this property shows how the simple addition and subtraction of word vectors can solve analogical reasoning tasks with remarkable accuracy. Consider the analogy *king:man :: queen:*______ ("king is to man as queen is to ?"). This question can be represented by the combination of word vectors $\overrightarrow{king} + \overrightarrow{woman} - \overrightarrow{man}$. Performing this operation and finding the closest word results in $\overrightarrow{queen}$. Figure 4-2 provides a graphical representation of this task. This figure illustrates how the relative positioning of these terms reflects a latent gender dimension captured by the vector $\overrightarrow{woman} - \overrightarrow{man}$ (Kozlowski, Taddy, and Evans 2019).

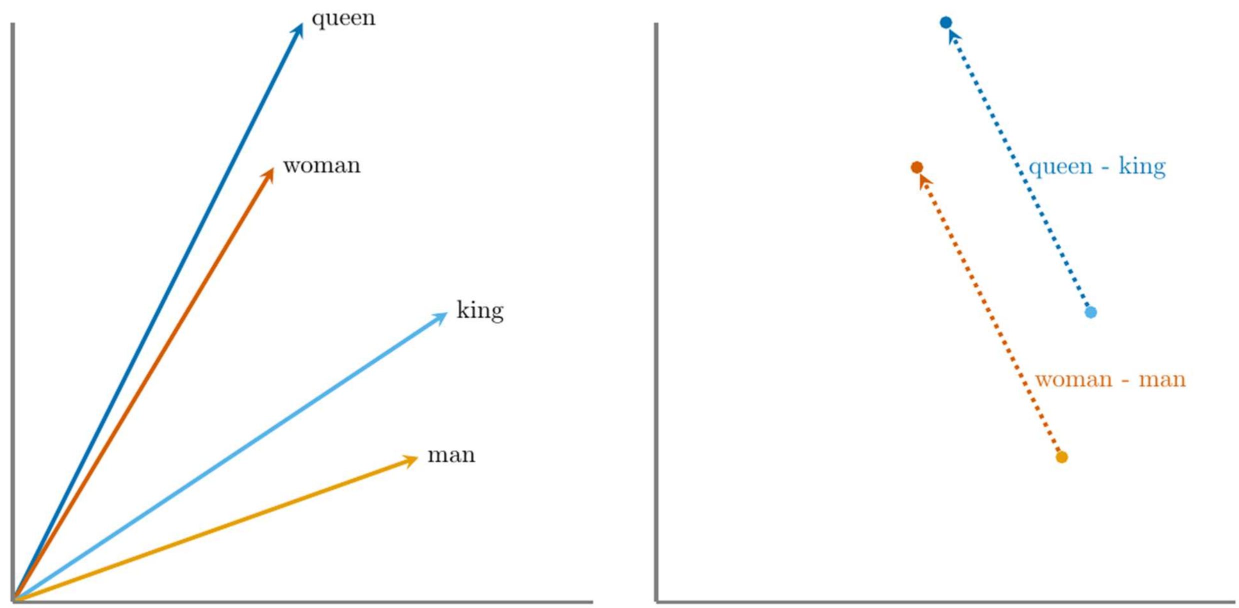

Figure 4-2. Example of a latent gender dimension in a word embedding model. The left panel shows the vector representations for the words king, queen, man, and woman. The right panel shows how the move from king to queen corresponds to the move from man to woman. The vector woman-man can be interpreted as moving along a "gender dimension." Figure adapted from *RasaHQ* (n.d.).

While *king* and *queen* are clearly gendered, word embedding models have also been shown to capture subtler but substantively important associations between concepts latent in text. For example, word embedding models trained on news articles have been shown to reflect occupational gender bias (Garg et al. 2018; Kozlowski, Taddy, and Evans 2019). Garg et al. (2018) measure occupational gender bias in word embeddings by first constructing vectors representing men and women. They do this by computing the average of lists of word vectors representing men (e.g., $\overrightarrow{men}, \overrightarrow{man}, \overrightarrow{boy}, \overrightarrow{he}, \overrightarrow{him}$) and women (e.g., $\overrightarrow{women}, \overrightarrow{woman}, \overrightarrow{girl}, \overrightarrow{she}, \overrightarrow{her}$). They then measure the Euclidean distance of these gender vectors to occupation vectors such as $\overrightarrow{doctor}$ and $\overrightarrow{engineer}$. The closer the occupation vector to average vector representing men, the more closely the occupation is associated with men, and similarly for women. Their measure of gender bias is then the difference distances to the gender vectors.

In the same way that gender-related terms can be used to construct a gender dimension, we can use other terms to construct a dimension relating to a theory of value. In the case of the theories of skill, I create a "skill" dimension by taking the average of the vectors $\overrightarrow{skill}$ and $\overrightarrow{skills}$. While it is possible to use more specific terms, this would require prior knowledge of the theories of value. A key advantage of the approach described here is that it can work even with more generic terms. Other terms that are closely associated with skill (in the sense that they frequently occur in similar contexts) in a particular corpus will thus be close to this vector. Cosine similarity is a commonly used similarity metric that measures the angle between two vectors and has the advantage that it is not affected by the magnitude of the vectors. The measure ranges from 1 to -1. A value of 1 indicates that the vectors are pointing in exactly the same direction and thus the two words are highly similar in that they appear in similar contexts. A value of -1 indicates that the vectors are pointing in exactly the opposite direction, indicating that the words are dissimilar and not used in the same context.

This approach can be readily extended to compare latent concepts across communities by simply training word embedding models on separate corpora for each community of interest. Comparing word embeddings trained on different corpora is commonly used to track changes in the meaning over words time, where the corpus is divided into multiple time slices and separate word embedding models are trained on each slice (e.g., Garg et al. 2018; Kozlowski, Taddy, and Evans 2019; Hamilton, Leskovec, and Jurafsky 2018). To identify elements comprising theories of skill in eSports and speedrunning, I collected discourse from two *Reddit* communities dedicated to these topics. I introduce these corpora below. I train separate word embedding models on corpora of discourse for each community and identify the terms closest to the skill vector. Examples of terms returned for eSports include mechanics, teamwork, and strategy, while terms returned for speedrunning include talent, dedication, and precision. These results are presented in full below.

To construct a coherence measure, one approach would be to identify the positions of the categories of interest relative to the vector of seed terms. For example, the positions of games in the eSports corpus could be measured relative to the skill vector. However, this approach will only work in the case that all categories of interest are mentioned in both corpora as no vector is defined for terms that do not appear in a corpus. Of the 26,712 unique games played by eSports and speedrun streamers in the *Twitch* data, 1,665 (6.2%) of these games were mentioned in the eSports corpus and 4,037 (15.1%) were mentioned in the speedrun corpus. However, the amount of time streamers spend playing games not mentioned in the corpora is relatively low—popular games like *Fortnite* and *League of Legends* account for a disproportionate amount of time spent streaming. Thus, the games that were mentioned in the corpora account for 67.4% and 50.4% of observations (streaming sessions) in the *Twitch* data, respectively. In order to measure the coherence of a larger percentage of these games, I construct a larger corpus consisting of discussions from additional general-purpose videogaming and livestreaming communities (described below). In this larger corpus, 18,705 games were mentioned, accounting for 70.0% of games and 95.1% of observations in the *Twitch* data.

While a word embedding model trained on the larger corpus has the advantage that vectors are defined for nearly all categories of interest, it has the notable disadvantage that the skill vector no longer represents the distinct theories of skill of the eSports and speedrunning communities. The skill vector in the larger corpus represents a kind of average concept of skill across all communities. Thus, simply measuring the relative positions of games to the skill vector is not representative of the coherence of games from the perspective of the eSports and speedrunning theories of skill. To address this issue, I use the terms that were previously identified as constituting theories of skill in the eSports and speedrunning communities. I then apply these terms to the larger corpus to derive vectors that represent eSports and speedrunning theories of skill. For example, the vector for eSports skill is the average of a list of word vectors including $\overrightarrow{mechanics}, \overrightarrow{teamwork}$, and $\overrightarrow{strategy}$, while the vector for speedrun skill includes $\overrightarrow{talent}, \overrightarrow{dedication}$, and $\overrightarrow{precision}$. The eSports and speedrun skills vectors thus represent the eSports and speedrunning theories of skill in the larger embedding space.

To measure the coherence of a pair of games, I take the difference in cosine similarity of the games to the newly defined skill vectors. I call this measure the relative cosine similarity. In the present case, relative cosine similarity can be measured in relation to the eSports and speedrun skill vectors, resulting in measures of eSports coherence and speedrun coherence. For example, to measure the eSports coherence of the games *Super Mario 64* and *Doom,* I calculate the cosine similarity of each game to the eSports skill vector in the larger corpus and then take their difference. This measure thus captures the extent to which two vectors (in this case, games) point in similar directions from the perspective of a third vector (in this case, the eSports or speedrun skill vector).

An important subtlety of this measure is that it does not measure whether a particular combination of offerings is considered a good application of a theory of value (i.e., whether or not the offerings are themselves valued). Such a measure would be maximized when each offering in a repertoire points in the same direction as the vector used to define the theory of value (i.e., the cosine similarity is 1 for each game in the repertoire). By contrast, the relative cosine similarity measure captures the extent to which the two vectors point in the same direction (when viewed from a particular perspective) and thus whether they might regarded as similar by someone applying a particular theory of value. An extreme case that illustrates this difference is two vectors that are identical and point in the opposite direction of the theory of value vector. These vectors would be considered extremely bad applications of that theory of value, but perfectly coherent because they point in the same direction.

Data: *Reddit* corpora

I now apply this approach to first identify the distinctive elements comprising eSports and speedrunning theories of skill, and second to develop measures of eSports coherence and speedrun coherence. As mentioned, I use three corpora. The eSports and speedrunning corpora consist of discussions from two prominent eSports and speedrunning communities. The "full" corpus which I use to measure the coherence of game pairs consists of both the eSports and speedrunning corpora as well as additional general-purpose communities related to videogames and livestreaming. To this end, I train word embedding models on discourse from two prominent communities dedicated to eSports and speedrunning, both hosted on *Reddit. Reddit* is a popular social media website consisting of many independent community forums, known as subreddits.

While there are numerous other Internet communities dedicated to eSports and speedrunning, the discussions found in the eSports and speedrun subreddits are useful for several reasons. First, both communities are prominent. The *eSports* subreddit was created in 2009 and consisted of 83.3 thousand members as of February 2022. The *speedrun* subreddit was created in 2011 and consisted of 201 thousand members as of February 2022. Second, The eSports and speedrun subreddits are for general discussion of topics related to eSports and speedrunning. By contrast, other prominent forums are dedicated to specific games. Finally, because both communities are on Reddit, this minimizes variation in text due to idiosyncratic factors and allows for a more direct comparison of the theories of skill used by these communities.

Error! Reference source not found. shows the documents comprising the eSports, speedrun, and full corpora. For each subreddit listed in the table, I collected all posts and submissions made before February 2022. After preprocessing the text, I trained separate word embedding models on the eSports and speedrunning corpora.***Error! Reference source not found.***

A key assumption for this approach is that at least some of the discussions in the *eSports* and *speedrun* subreddits must reflect latent theories of skill. Particularly in the case of the kinds of online communities being studied here, community forums are a primary channel for generating, disseminating, and negotiating theories of value. Reading hundreds of posts made to these communities, this assumption appears to be satisfied.

While many posts reflect theories of skill implicitly, the eSports and speedrun communities each in fact frequently engage in explicit discussions about skill. For example, in a thread about which games have the highest mechanical and strategic skill ceiling, one user on the eSports subreddit contrasts the kinds of skill required to compete at a high level in two popular fighting games (*Melee has the hardest learning curve*, 2017):

[Super Smash Brothers:] Melee has the hardest learning curve as far as advanced techniques and require extreme precision to execute especially in the heat of a match. Street Fighter becomes harder once you play on a higher level. It's less about the inputs and more of the mental fortitude required to beat your opponent.

Table 4-1. Characteristics of documents in the *Reddit* corpus. The full corpus uses posts from all 23 subreddits. The eSports and speedrun corpora use only posts from those respective subreddits.

Subreddit	Total posts	Unique authors	First post
gaming	47,921,115	4,143,096	2007-09-17
Games	16,361,561	646,851	2008-08-08
NintendoSwitch	10,385,244	775,396	2016-10-20
xboxone	9,466,512	676,706	2013-05-21
PS4	8,501,111	732,950	2010-05-17
pcgaming	7,095,481	557,454	2008-03-19
LivestreamFail	5,206,868	424,553	2016-01-14
nintendo	3,027,519	379,318	2008-10-28
Steam	2,802,326	468,465	2009-12-29
Twitch	2,051,278	348,243	2012-09-14
truegaming	1,871,833	143,205	2011-05-03
gamingsuggestions	1,103,762	130,118	2011-10-14
letsplay	874,807	49,968	2010-08-19
gamernews	697,919	101,335	2011-02-21
speedrun	592,724	76,812	2011-07-24
IndieGaming	491,471	90,112	2009-06-15
esports	190,472	41,833	2009-06-15
twitchstreams	174,191	40,486	2013-10-27
streaming	114,368	33,817	2009-07-21
mixer	76,773	13,172	2017-07-01
youtubegaming	50,814	14,324	2012-05-06
LivestreamFails	39,273	12,693	2014-04-29
Total	119,097,422	6,299,268	-

Explicit discussions about skill are also present in the *speedrun* subreddit. For example, in response to a question about which games and categories take the most skill to speedrun, *Reddit* user 6000j (2021) replies:

> what kind of skill? Mechanical skill? Memorisation? Quickly adapting to various situations? Even within each of those categories, there's a huge diversity, the mechanics to speedrun a 3d game tend to be pretty dang different from a 2d game, and so on.

Results

Figure 4-3 uses a tanglegram to compare the 25 terms most closely associated with skill in the eSports and speedrun corpora. This figure shows the relative importance of terms, as measured by their cosine similarity to the skill vector, as well as the correspondence between terms across the two corpora. By showing the relative importance of the terms in the two corpora, the tanglegram provides a useful way of visually comparing contextual differences in latent concepts like skill. The more tangled the lines, the lower the correspondence between the two models. To my knowledge, this is the first use of tanglegrams to compare word embeddings. **Error! Reference source not found.** provides further information about the method and compares it to other common methods of comparing word embedding models.

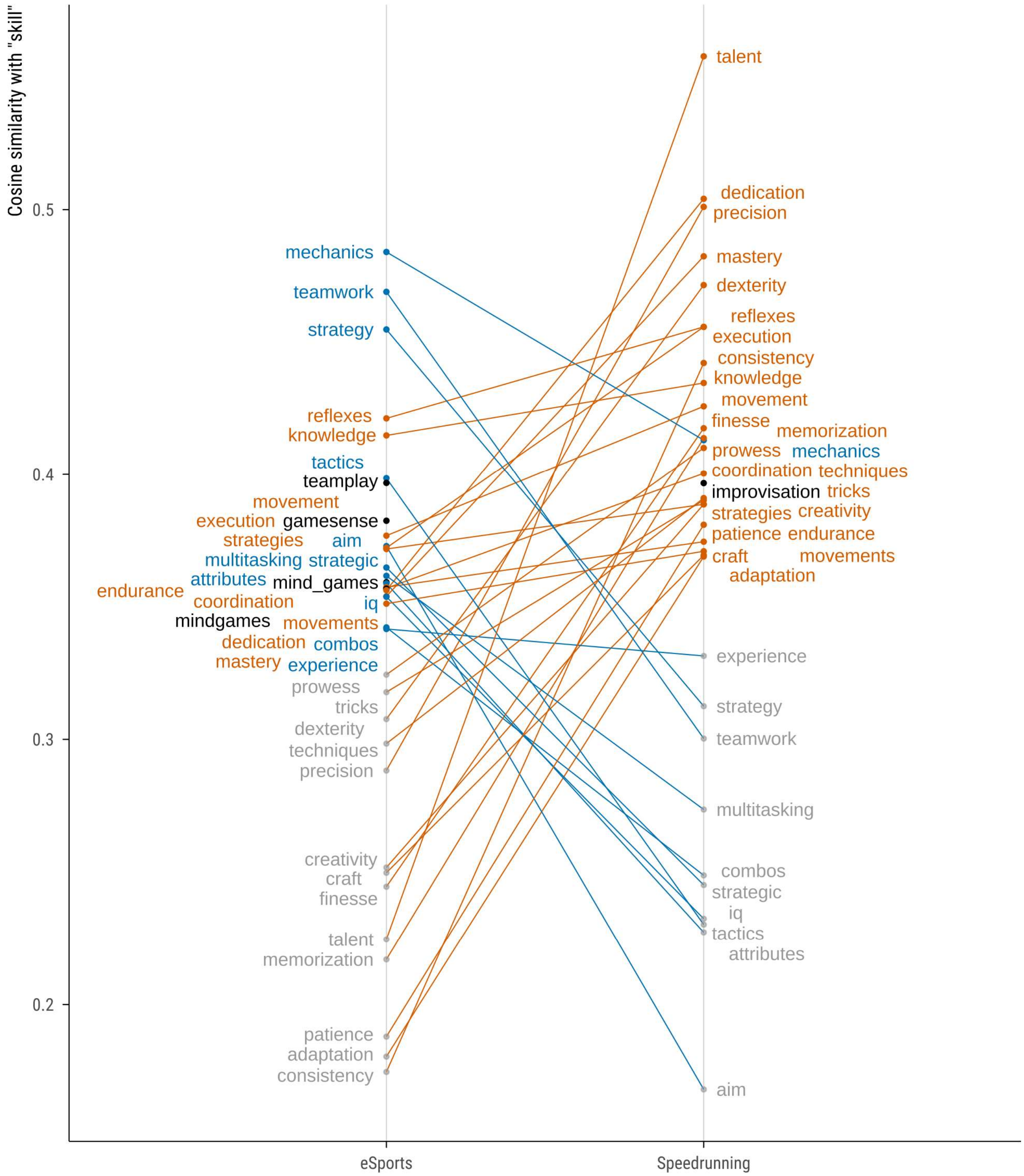

Figure 4-3. Tanglegram comparing eSports and speedrunning theories of skill. The y-axis represents cosine similarity of a term with the vector `skill + skills` within a community. Blue lines indicate skills that rank higher for eSports than speedrunning; orange lines indicate skills that rank higher for speedrunning than eSports. Terms in black do not appear in the other community's vocabulary and thus do not have a corresponding similarity score. Terms written in gray are not in the top 25 of a dimension. Synonyms for "skill" were omitted from the results.

Overall, these descriptive results reveal a pattern of partially overlapping but distinct theories of value that is consistent with my informal observations of the eSports and speedrunning communities. 11 terms are in the top 25 of both communities. Some degree of overlap is expected as many videogames played by both eSports and speedrun streamers require general skills such as mechanics[13], reflexes, and knowledge. Within these overlapping terms, however, their relative importance can vary significantly within eSports and speedrunning. For example, "mechanics" receives the highest score in eSports, but is ranked 13th in speedrunning.

At the same time, many terms are unique to the top 25 of the respective community and in some cases do not appear in the other community. In the case of eSports, many of the unique terms reflect the team-based and multiplayer nature of the games. For example, "teamwork" is the second-highest skill in eSports, but is ranked 28th in speedrunning. Similarly, the terms "teamplay," "gamesense", and "mindgames"[14] are in the top 20 terms in eSports but these phrases do not appear in the speedrun community corpus.

As discussed previously, this approach was designed to identify dimensions of shared understanding and not necessarily dimensions on which all members of the community agree. Based on my qualitative data acquired through fieldwork and informal observations, most of the skills identified by the word embedding model are ones in which community members are likely

[13] In the context of videogames, "mechanics" broadly refers to the "the rules, processes, and data at the heart of a game. They define how play progresses, what happens when, and what conditions determine victory or defeat… Expert players … will learn how to use a game's core mechanics to their advantage" (Adams and Dormans 2012, Ch. 1). The types of mechanics that expert players need to perfect can vary greatly across games. In shooting games, core mechanics often include "detailed physics for movement, shooting, [and] jumping," whereas in strategy games core mechanics often include resource management and unit positioning (Adams and Dormans 2012, Ch. 1). Mechanics that are useful in eSports are often not useful in speedrunning, and vice versa. In particular, speedrunners often exploit glitches in games, which would be not be allowed in eSports.

[14] Game sense and mind games refer to psychological aspects of competitive gameplay. Game sense refers to a player's instincts around what is occurring in the game, especially with regard to their teammates or opponents. Mind games are acts of "calculated psychological manipulation, done especially to confuse" an opponent and gain strategic advantage (American Heritage Dictionary, 2011).

to be in consensus. However, the word embedding model also identifies skills that are more controversial. In particular, “rng” (random number generator) and “improvisation” are polarizing terms within the speedrunning community.[15] These terms refer to situations where the game introduces random elements that require on-the-fly adaptation. Some members of the speedrun community find such games more entertaining because of the variety and also regard the ability to adapt to randomness a distinct skill. On the other extreme are games where each run is identical. Other members of the speedrunning community prefer deterministic games because they provide a level playing field (randomness in games can change the time it takes to complete the game) and allow for greater optimization, to the point where some players complete the games blindfolded.

In short, these descriptive results are consistent with my qualitative data and informal observations, validating a key assumption of the data and empirical approach, which is that latent theories of skill are reflected in community discourse and the word embedding model is able to detect subtle differences in these theories of skill across communities. I provide further validation of the word embedding approach in the main results section.

As described in the methods section, these results were used to generate vectors representing eSports skill and speedrunning skill in the full *Reddit* gaming corpus. These vectors

[15] Controversy surrounding the role of randomness in speedrunning is evident in community discussions. To give one example, *Reddit* user Soulcloset (2019) writes in the speedrun subreddit:

> In many games, the RNG [random number generation] and adapting to it is part of what makes the game fun. As much as runners complain about RNG, there are instances in which getting rid of it would remove some of the appeal of the run.
>
> For example, I run Barbie Horse Adventures: Wild Horse Rescue. In that game, there are small animals which run in a sometimes random pattern, getting in the way of your horse and attacking it. If these were completely predictable, the run would be solely based on execution, and given that it's fairly lenient, the time wouldn't come down nearly as much as it has.

were then used to create measures of eSports coherence and speedrun coherence, capturing the extent to which a pair of games appears similar from the perspective of someone applying one of these theories of skill. I return to these measures and validate them in the next section.

REPERTOIRE COHERENCE AND AUDIENCE VALUATION

Having established a method for measuring theoretical coherence, I now use data on streamers' game repertoires and their viewership to test whether audiences penalize incoherence. These data consist of all streams by eSports and speedrun streamers that occurred on *Twitch* between October 2019 and February 2022. To explore how theories of value imply different but consequential ways of organizing offerings and thereby which combinations of categories are viable, I focus my analyses on predicting the differential changes in viewership for eSports and speedrun streamers associated with eSports and speedrun incoherence.

Insofar as eSports and speedrun streamers are evaluated according to different theories of skill, I make two basic predictions:

> **Hypothesis 1** (incoherence penalty): The greater the eSports (speedrun) incoherence of a game relative to the rest of an eSports (speedrun) streamer's repertoire, the lower their viewership.

> **Hypothesis 2** (type incoherence): eSports (speedrun) streamers are penalized more for eSports (speedrun) incoherence than speedrun (eSports) incoherence.

Given the prominence of typicality-based explanations of audience valuations, a natural question arises regarding the relationship between coherence and typicality. To explore this relationship further, I include a measure of typicality in my analyses. As the primary purpose of this paper is to develop and test the concept of theoretical coherence, I have not theorized this relationship explicitly and therefore make no ex ante predictions. Nevertheless, I use these analyses as an opportunity to further explore the relationship between coherence and typicality. In particular, I consider whether coherence continues to have an effect after controlling for typicality.

My unit of analysis is the game-session, which I define as an interrupted period of time playing a single game. A single streaming session can thus result in multiple observations. For example, a streamer who begins their stream by playing Fortnite, switches to "Just Chatting," and then switches back to Fortnite before logging off would have three observations despite all the game-sessions occurring within a single stream.

MODEL

The model predicts logged average viewers of streamer *i* in game-session *t*:

$$\begin{aligned}
\log(avg_viewers_{it}) \\
&= is_esports_i \times \boldsymbol{\beta}_1 \boldsymbol{X}_{it} \\
&+ is_speedrunner_i \times \boldsymbol{\beta}_2 \boldsymbol{X}_{it} \\
&+ \boldsymbol{\beta}_3 \boldsymbol{C}_{it} + \alpha_i + \gamma_{it} + \lambda_{it} + \epsilon_{it},
\end{aligned}$$

where $\boldsymbol{X}_{it}$ represents a vector of the main independent variables:

$$[esports\_incoherence_{it}, speedrun\_incoherence_{it}, typicality_{it}]^T.$$

To study the differential effects of coherence on eSports and speedrun streamers, I interact the main independent variables with dummy variables indicating whether the streamer is an eSports streamer $is\_esports_i$ or a speedrun streamer $is\_speedrunner_i$. Streamers in the dataset are only ever of one type and do not change type. Consequently, $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ represent the effects of the main independent variables for eSports and speedrun streamers, respectively. $\boldsymbol{C}_{it}$ represents a vector of time-varying streamer level controls discussed below. α_i are streamer fixed effects, γ_{it} are fixed effects for the date the game-session the game session was started, λ_{it} are game fixed effects, and ϵ_{it} is an error term. To account for heteroskedasticity and serial correlation, I cluster the errors at the streamer, date, and game level.

SAMPLE INCLUSION CRITERIA

I used two additional data sources to identify eSports and speedrun streamers to include in my sample: player profiles on *Liquipedia* and *Speedrun.com. Liquipedia* and *Speedrun.com* are popular websites that catalog players who compete in eSports and speedrunning, respectively. Players who are featured on these sites are thus recognized as bona fide eSports competitors or speedrunners by their respective communities. Additional information about these data sources and how the samples were constructed is provided in **Error! Reference source not found.**.

After identifying these streamers, I further limited my sample to address two concerns. First, many streamers have virtually no viewers. Audience valuations of these streamers cannot be estimated because they have no meaningful audience. I thus excluded streamers who: streamed less than 10 times, streamed less than 10 hours total, or had an average viewership of less than one.

Second, it is possible that some streamers who appear in the *Liquipedia* and *Speedrun.com* data do not engage in eSports or speedrunning on their Twitch streams. For example, a retired eSports competitor may engage in variety streaming. Some variety streamers might be included in the *Speedrun.com* data because they occasionally engage in speedrunning. Variety streamers as assumed to be subject to a different theory of value emphasizing charisma and close interaction with the audience. I thus excluded streamers who spent more than 50% of the time in "Just Chatting."

I also excluded streamers who spent more than 25% of the time in uncategorized streams. Streams are occasionally uncategorized streams because the streamer forgets to set the game they are playing or when they are transitioning between games. Streamers have a strong incentive to (accurately) categorize their streams because it promotes discoverability. Streamers who spend a significant portion of time as uncategorized is thus unlikely to be strategic and responsive to audience valuations.

Finally, after calculating the measures detailed below, I restricted my sample to streams occurring on or after January 1, 2021. This was done to allow cumulative measures to stabilize for streamers who began streaming prior to the observation period.

MEASURES

The main dependent variable is the logged average concurrent viewers in a game-session. Concurrent viewers refer to the number of viewers a streamer has at a given point in time. Average concurrent viewership is calculated by taking the mean concurrent viewers of each observation comprising a game session.

Independent variables

*eSports and speedrun incoherence.—*The main independent variables are eSports incoherence and speedrun incoherence. These variables capture the extent to which the current game coheres with the streamer's existing repertoire with regard to the eSports and speedrun theories of skill. They are calculated by taking the absolute distance of the current game from the repertoire centroid on the eSports and speedrunning dimensions defined in the previous section. The repertoire centroid of streamer *i* in game-session *t* is a weighted average of the eSports or speedrun score $score_g$ of all prior games in the streamer's repertoire,

$$repertoire\ centroid_{it} = \sum_g weight_{gi(t-1)} \cdot score_g,$$

where $weight_{gi(t-1)}$ is the proportion of time spent streaming game *g.*[16]

The extent to which incoherence affects viewership can be expected to vary by the prior level of incoherence in the streamer's repertoire. A streamer who consistently plays highly varied games will have attracted viewers who have a greater tolerance for incoherence. To account for this, I normalize the incoherence measure by the standard deviation of the repertoire.[17] This

[16] The proportion of time spent streaming game *g* is the cumulative number of hours *h* streaming game *g* divided by the cumulative number of hours streaming any game: $weight_{gi(t-1)} = \frac{hours\ streamed_{gi(t-1)}}{\sum_g hours\ streamed_{gi(t-1)}}$.

[17] One drawback to normalizing the distance from the centroid is that it does not handle concentrated repertoires well. In the case where the streamers prior repertoire consists of a single game, the standard deviation of the repertoire will be 0, and thus the normalized distance is undefined. Streamers with highly concentrated repertoires have very small standard deviations, resulting in an unreasonable adjustment that over is that distance is undefined when the streamer has been playing only one game and adds a second game. A second drawback is that streamers with a highly concentrated portfolio will have very small standard deviations, resulting in an overly large normalized distance. The former problem occurs at most once per streamer, and thus I disregard it as having a negligible impact on the final outcome. I address the latter problem by censoring the variable so that the minimum and maximum values are -+10.

standard deviation is also weighted by time (the formula for a weighted standard deviation is provided by National Institute of Standards and Technology 1996):

$$repertoire\ standard\ deviation_{it} = \sqrt{\frac{\sum_g weight_{gi(t-1)} \left(score_g - repertoire\ centroid_{it}\right)^2}{\frac{N-1}{N}}},$$

Where N is the number of games in the repertoire. The normalized incoherence of the current game g^* with respect to the streamer's repertoire is thus

$$incoherence(game, repertoire_{it}) = \frac{\left|score_g - repertoire\ centroid_{it}\right|}{repertoire\ standard\ deviation_{it}}.$$

Because the eSports and speedrun theories of skill only apply to videogames, I exclude any sessions where the streamer is in Just Chatting. I also exclude any uncategorized sessions as no score can be calculated for these sessions.

Typicality.—I also include a measure of typicality. Following prior work (e.g., Goldberg, Hannan, and Kovacs 2016), I measure typicality using a Jaccard similarity score. For a pair of games g_1 and g_2, the Jaccard score is the number of repertoires containing both games divided by the number of repertoires containing either game[18]

$$jaccard(g_1, g_2) = \frac{|g_1 \cap g_2|}{|g_1 \cup g_2|}.$$

[18] In the event $g_1 = g_2$, I set the Jaccard score to 1.

The score ranges from 0 when the games never co-occur in a repertoire to 1 when the two games always co-occur. Like the incoherence measures, I measure the typicality of a game with respect to a repertoire using a weighted average of the typicality of the current game g^* with all games previously appearing in the streamer's repertoire,

$$typicality(g^*, repertoire_{it}) = \sum_g weight_{gi(t-1)} \cdot jaccard(g^*, g).$$

Time-varying controls

I include controls for several time-varying variables that could confound the relationship between incoherence and viewership:

- Repertoire concentration: A measure of repertoire concentration based on the Herfindahl-Hirschman Index $HHI_{it} = \sum_g s_{gi(t-1)}^2$ where s_{git} is the proportion of time spent streaming game *g* by streamer *i* as of session *t.* Time spent in "Just Chatting" and uncategorized streams is not included in the calculation.
- Cumulative hours streaming (log): the logged cumulative number of hours the streamer has spent streaming games. Time spent in "Just Chatting" and uncategorized streams is not included in the calculation.
- Cumulative number of unique games (log): the logged cumulative number of unique games in the streamer's repertoire, not including the current game session. "Just Chatting" and uncategorized streams are not included in the calculation.
- Cumulative hours in "Just Chatting" (log): the logged cumulative number of hours the streamer has spent streaming in the "Just Chatting" category, not including the current game session. I include this as a proxy for relational labor. In the second chapter of this

dissertation, I find that viewers are more tolerant of experimentation with new games when the streamer engages in more relational labor.

- New stream: A dummy variable that takes 1 if the game session occurs at the beginning of a new stream. Streamers
- First time playing game: A dummy variable that takes 1 the first time a streamer plays a game.
- Single-game repertoire: A dummy variable that takes 1 if the streamer only has one game in their repertoire.

Descriptive statistics are provided in Tables

Table 4-2 and

Table 4-3. Correlations are shown in

Table 4-4.

Table 4-2. Stream-game session descriptive statistics, by streamer type.

	eSports (n=2,228,913)				Speedrun (n=3,626,900)			
	Min	**Max**	**Mean**	**SD**	**Min**	**Max**	**Mean**	**SD**
Avg. concurrent viewers	0.00	443396.08	672.15	3305.09	0.00	41230.79	39.46	285.26
eSports incoherence	0.00	5.00	0.68	1.08	0.00	5.00	0.78	0.83
Speedrun incoherence	0.00	5.00	0.69	1.10	0.00	5.00	0.79	0.83
Typicality	0.00	1.00	0.68	0.35	0.00	1.00	0.36	0.34
Repertoire concentration	0.01	1.00	0.69	0.28	0.01	1.00	0.38	0.29
Cum. hours streaming	0.00	18555.52	1111.23	1239.50	0.00	9933.89	684.78	788.95
Cum. hours streaming game	0.00	17032.30	564.85	815.33	0.00	8354.25	167.91	326.02
Cum. hours in Just Chatting	0.00	2657.92	25.90	101.68	0.00	1968.04	10.99	57.00
Cum. hours uncategorized	0.00	849.14	3.39	27.22	0.00	762.18	5.12	22.52
Cum. number of game sessions	1	2023	173.16	201.04	1	1563	67.34	109.09
Cum. number of unique games	0	540	17.98	33.50	0	521	28.97	36.17
New stream	0	1	0.17	0.37	0	1	0.17	0.38
Changed game from prev. session	0	1	0.32	0.47	0	1	0.47	0.50
First time playing game	0	1	0.04	0.19	0	1	0.09	0.29
Single-game repertoire	0	1	0.09	0.28	0	1	0.04	0.19

Table 4-3. Repertoire descriptive statistics, by streamer type.

	eSports (n=8,742)				Speedrun (n=15,362)			
	Min	**Max**	**Mean**	**SD**	**Min**	**Max**	**Mean**	**SD**
Stream sessions	0	4495	42.90	133.69	0	2511	40.43	98.40
Stream-game sessions	10	5786	275.38	347.23	10	3680	252.84	278.16
Repertoire concentration	0.01	1.00	0.74	0.26	0.01	1.00	0.39	0.29
eSports skill centroid	-0.10	0.32	0.19	0.09	-0.24	0.31	-0.01	0.07
Speedrun skill centroid	-0.28	0.14	-0.16	0.08	-0.26	0.31	0.02	0.06
eSports skill std. dev.	0.00	0.25	0.05	0.04	0.00	0.27	0.06	0.04
Speedrun skill std. dev.	0.00	0.25	0.04	0.03	0.00	0.27	0.06	0.03
Cum. hours streaming	10.11	18555.52	772.88	1164.92	10.00	9933.89	572.73	765.66
Prop. time Just Chatting	0.00	0.49	0.01	0.03	0.00	0.50	0.01	0.03
Cum. number of unique games	1	540	11.74	21.21	1	521	25.48	33.28
Single-game repertoire	0	1	0.13	0.34	0	1	0.03	0.18

Table 4-4. Correlation table. Variables are at the game-session level.

		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
1.	Avg. concurrent viewers	1	.	.	.	.	.	.	.	.	.	.	.	.	.
2.	eSports incoherence	.00	1	.	.	.	.	.	.	.	.	.	.	.	.
3.	Speedrun incoherence	.00	.91	1	.	.	.	.	.	.	.	.	.	.	.
4.	Typicality	.02	-.55	-.56	1	.	.	.	.	.	.	.	.	.	.
5.	Repertoire concentration	-.01	-.19	-.19	.79	1	.	.	.	.	.	.	.	.	.
6.	Cum. hours streaming	.18	.06	.07	-.09	-.13	1	.	.	.	.	.	.	.	.
7.	Cum. hours streaming game	.10	-.25	-.25	.45	.32	.57	1	.	.	.	.	.	.	.
8.	Cum. hours in Just Chatting	.26	.05	.06	-.11	-.15	.36	.06	1	.	.	.	.	.	.
9.	Cum. hours uncategorized	.02	.02	.02	-.08	-.10	.17	.00	.17	1	.	.	.	.	.
10.	Cum. number of game sessions	.11	-.27	-.27	.53	.34	.49	.87	.16	.03	1	.	.	.	.
11.	Cum. number of unique games	.23	.08	.08	-.43	-.53	.50	-.04	.35	.25	-.01	1	.	.	.
12.	First time playing game	-.01	.19	.19	-.34	-.14	-.05	-.15	-.01	.00	-.19	.10	1	.	.
13.	Changed game from prev. session	.03	.30	.31	-.51	-.40	.13	-.20	.12	.05	-.22	.29	.33	1	.
14.	New stream	.07	.23	.24	-.31	-.20	.16	-.12	.16	.04	-.12	.22	.16	.53	1
15.	Single-game repertoire	-.02	-.20	-.20	.34	.38	-.16	-.06	-.05	-.02	-.05	-.16	-.04	-.20	-.09

VALIDATING COHERENCE MEASURES

I conduct two basic tests to validate the eSports and speedrun scores and coherence measures. Insofar as streamers generally play games that are consistent with their type and construct repertoires that are coherent according to their type, we should expect (1) that eSports/speedrun streamers' repertoire score higher on the eSports/speedrun skill dimension, and (2) that eSports/speedrun streamers play games that are closer on the eSports/speedrun dimension than on the speedrun/eSports dimension.

Figure 4-4 shows the distribution of repertoire centroids for eSports and speedrun streamers. As expected, streamers construct repertoires of games that are more closely associated with the theory of skill corresponding to their type. Figure 4-5 plots the incoherence measure for eSports and speedrun streamers.

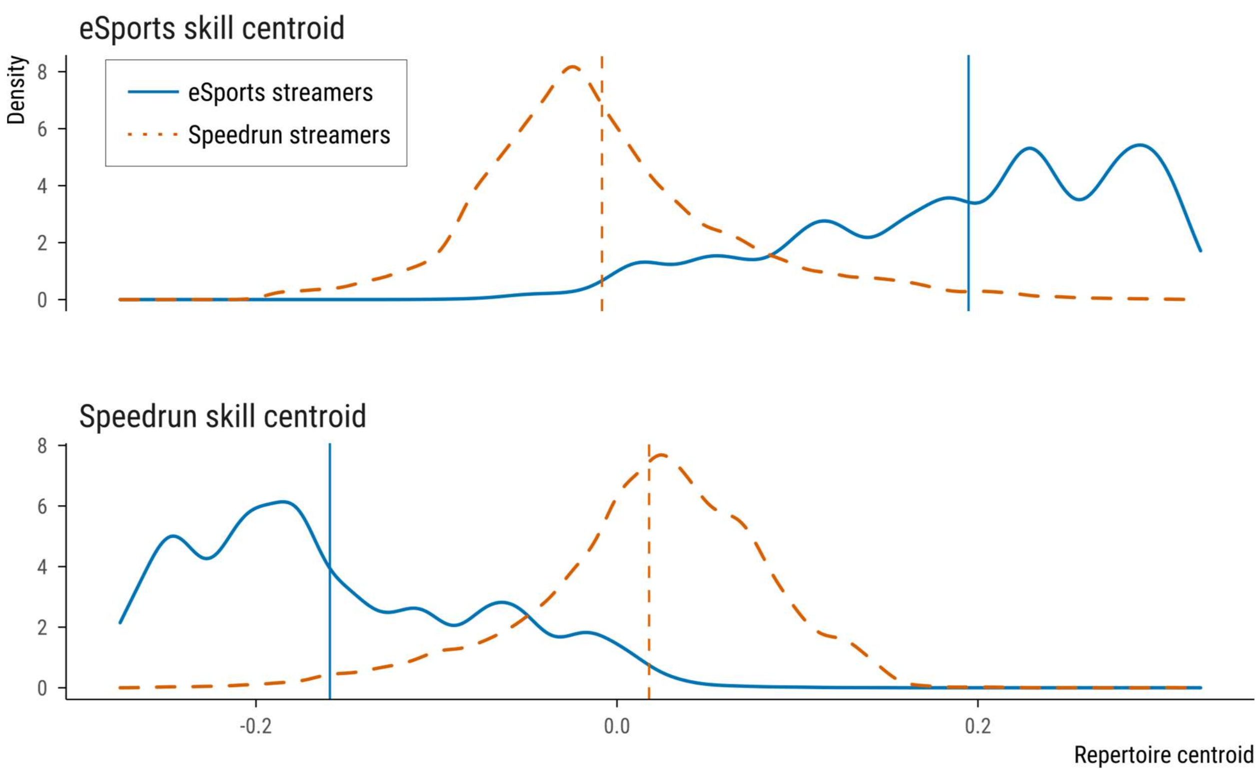

Figure 4-4. Distribution of repertoire centroids, by streamer type.

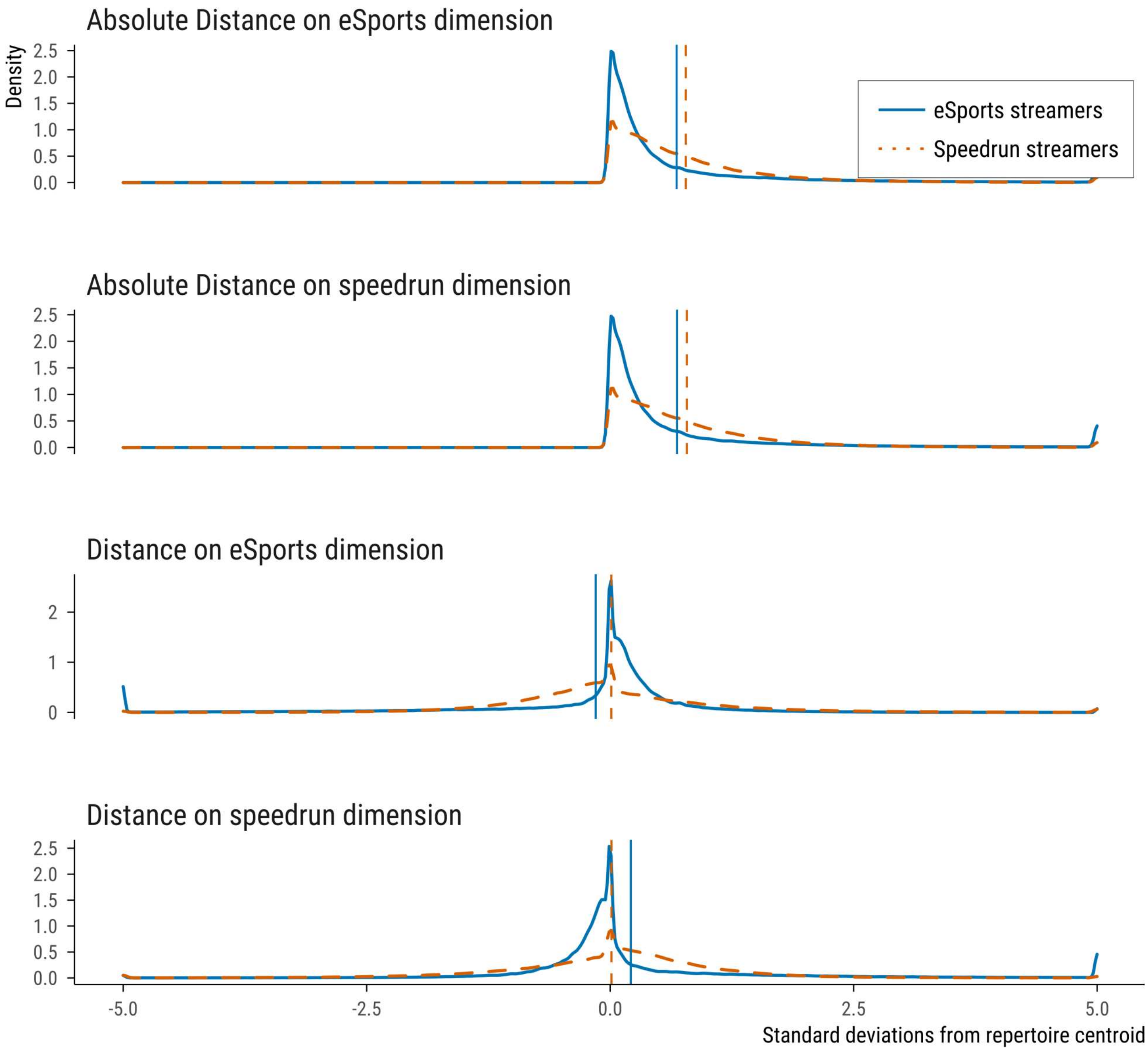

Figure 4-5. Distribution of game distance from repertoire centroid, by streamer type. Distances are normalized by the standard deviation of the streamer's repertoire centroid.

MAIN RESULTS

Table 4-5 reports the main results. Models 1-6 provide baseline estimates of the different combinations of incoherence, typicality, and time-varying controls. Model 7 presents the fully specified model. To allow for easier interpretation of the main independent variables,

Table 4-5. Fixed effects regression results predicting average viewership in a game session. The number of observations varies because incoherence scores could not be calculated for games not mentioned in the *Reddit* corpus.

	Avg. viewer count (log)						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
eSports streamer × ...							
eSports incoherence	-0.087***		-0.014*		-0.055***		-0.020***
	(0.020)		(0.008)		(0.012)		(0.007)
Speedrun incoherence	-0.027**		0.039***		-0.016**		0.024***
	(0.011)		(0.008)		(0.008)		(0.007)
Typicality		0.895***	0.957***			0.617***	0.619***
		(0.053)	(0.064)			(0.037)	(0.039)
Speedrun streamer × ...							
eSports incoherence	-0.023***		0.003		0.0003		-0.003
	(0.004)		(0.004)		(0.005)		(0.004)
Speedrun incoherence	-0.036***		-0.005		-0.006		-0.012***
	(0.005)		(0.005)		(0.005)		(0.005)
Typicality		0.338***	0.341***			0.062***	0.033
		(0.014)	(0.020)			(0.024)	(0.027)
Time-varying controls							
Repertoire concentration				0.474***	0.505***	0.388***	0.405***
				(0.084)	(0.090)	(0.087)	(0.090)
Cum. hours streaming (log)				0.233***	0.243***	0.247***	0.248***
				(0.020)	(0.021)	(0.020)	(0.021)
Cum. hours streaming game (log)				0.115***	0.104***	0.083***	0.083***
				(0.007)	(0.006)	(0.004)	(0.004)
Cum. number of unique games (log)				-0.032	-0.036*	-0.035*	-0.037*
				(0.020)	(0.021)	(0.021)	(0.021)
Cum. hours in Just Chatting (log)				0.015	0.015	0.028***	0.027***
				(0.011)	(0.011)	(0.008)	(0.008)
New stream=1				0.266***	0.271***	0.251***	0.251***
				(0.024)	(0.025)	(0.006)	(0.007)
First time playing game=1				0.173***	0.163***	0.157***	0.157***
				(0.016)	(0.015)	(0.007)	(0.007)
Single-game repertoire=1				-0.064***	-0.087***	-0.104***	-0.107***
				(0.023)	(0.024)	(0.022)	(0.023)
# Streamer fixed effects	18,278	25,200	18,221	25,202	18,223	25,200	18,221
# Game fixed effects	24,099	24,104	24,097	24,104	24,099	24,104	24,097
# Day fixed effects	767	767	767	767	767	767	767
Observations	5,548,020	5,612,108	5,347,989	5,841,568	5,533,775	5,612,108	5,347,989
R^2	0.891	0.892	0.893	0.898	0.899	0.899	0.899
Within R^2	0.010	0.033	0.033	0.081	0.084	0.091	0.091

All models include fixed effects for streamer, game, and day.
Standard errors are in parentheses and clustered by streamer, game, and day.
+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

The results show that streamers are penalized when they play games that do not cohere with the theory of skill associated with their type. Compared to playing a perfectly coherent game, model 7 predicts that eSports streamers will experience 2.0% (p<.001) fewer concurrent viewers and speedrun streamers 1.2% (p<.001) fewer concurrent viewers when they play games that are one standard deviation from their repertoire centroid with respect to their type.[19] For both eSports and speedrun streamers, the penalty for incoherence with regard to their type is minimized when incoherence is 0.

Furthermore, incoherence with regard to the streamer's type is penalized more severely than incoherence with regard to other theories of skill. For eSports streamers, model 7 predicts a 2.4% (p<.01) increase in concurrent viewers when playing a game that is one standard deviation from the repertoire centroid on the speedrunning dimension. For speedrun streamers, model 7 predicts a 0.3% (n.s.) decrease in concurrent viewers when playing a game that is one standard deviation from the repertoire on the eSports dimension, although this effect is not statistically significant.

The statistically significant negative association between viewership and incoherence with regard to the streamer's type as well as the more severe penalty for incoherence with regard to the streamer's type are consistent across the alternative specifications in models 1-7 are robust to controlling for typicality and time-varying controls.

Although not part of a formal hypothesis test, it is interesting to note that controlling for typicality results in a substantively large change in the coefficient for coherence. Comparing predicted values from models 5 and 7 for eSports streamers, controlling for typicality reduces the effect of a one standard deviation increase in incoherence (with respect to their type) from -5.4% to -2.0%. For speedrun streamers, however, controlling for typicality increases the penalty for incoherence from -0.6% (n.s.) to -1.2%. Furthermore, the substantive effect of typicality is much larger than that of incoherence for eSports streamers, but not statistically significant for speedrun

[19] Percentage change in viewership for eSports streamers associated with a 1 unit increase in eSports incoherence was calculated as $100(e^{-0.02} - 1) = -2.0$.

streamers. These results further reflect the differential (correlational) effects of typicality on eSports and speedrun streamers.

To what extent is theoretical coherence able to explain variation in audience valuations that prototype theory cannot? On the one hand, coherence continues to be significantly associated with valuations even after controlling for typicality in model 7. This suggests that coherence plays an independent role in valuation above and beyond typicality. Valuation of category spanning offerings cannot therefore be merely reduced to typicality. On the other hand, introducing typicality into the model results in a substantively large decrease in the predicted effect of coherence on viewership. Furthermore, the predicted effect on viewership is much greater for typicality than coherence. To evaluate the significance of these findings and their implications for prototype theory and the theory of value approach requires clarification of the relationship between coherence and typicality. I now turn to a post hoc explanation of this relationship, which raises several challenging questions that can be fruitfully addressed by future research.

THE RELATIONSHIP BETWEEN COHERENCE AND TYPICALITY

The preceding results suggest a close relationship between coherence and typicality. The relationship is plotted in Figure 4-6. How should we regard this relationship and interpret this result? As a first step toward unpacking this relationship and potentially synthesizing prototype theory and the theory of value approach, I first consider how these results can be interpreted as support for a weak version of prototype theory. I then consider the extent to which coherence and typicality represent different concerns that audiences might have.

Strong vs. weak forms of prototype theory

One interpretation of this result is that prototype effects take primacy over coherence in audience valuations. On the surface, this interpretation appears reasonable because of the relatively large

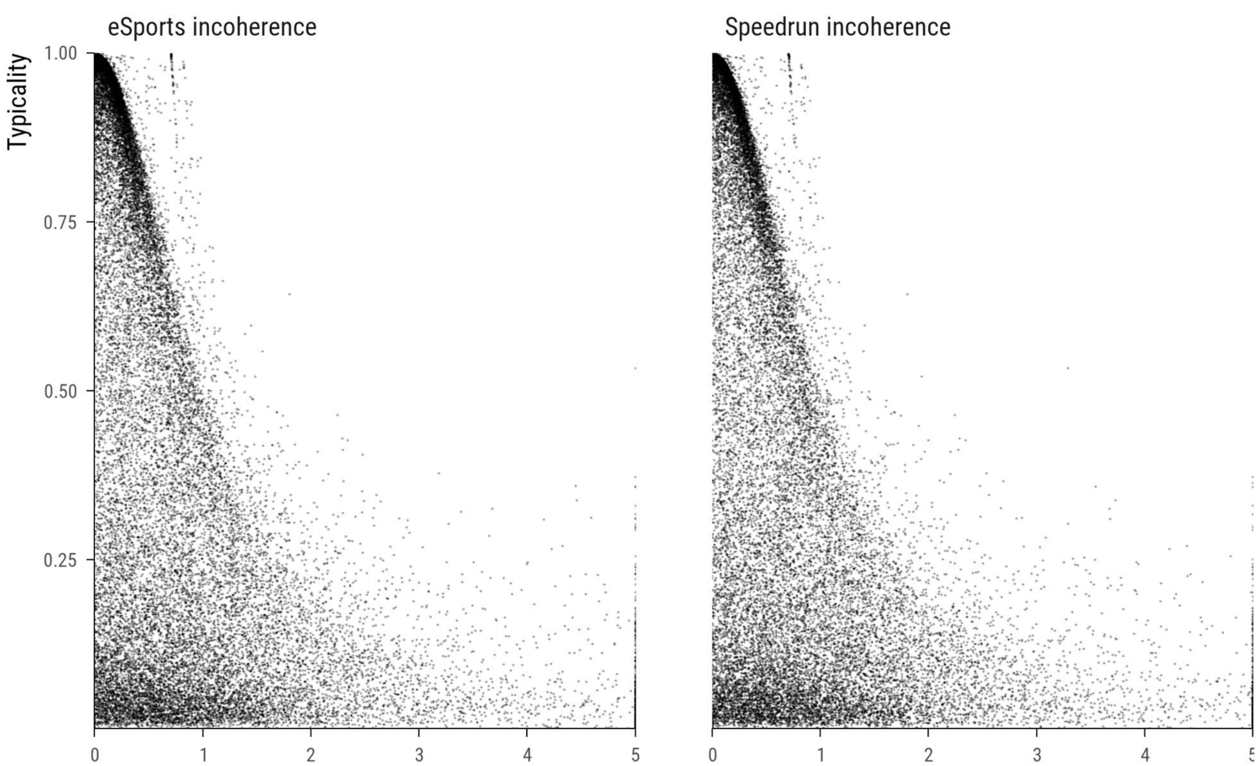

Figure 4-6. Relationship between incoherence and typicality. Based on a 0.5% random sample.

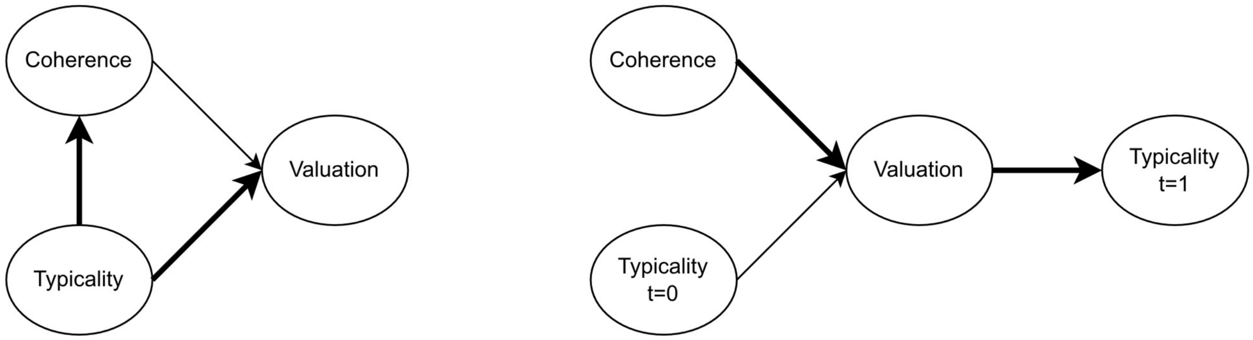

Figure 4-7. Causal pathways implied by strong (left) and weak (right) versions of prototype theory. Heavy arrows indicate larger effects. In the strong version of prototype theory, theories of value are determined by cognitive limitations in our ability to process and represent category-spanning offerings. In the weak version of prototype theory, prototypes emerge and are highly correlated with value, but are the result of dominant theories of value.

associations between typicality and viewership compared to coherence. However, this interpretation requires strong assumptions about the causal relationship between typicality and coherence. To clarify the assumed causal relationships, prototype theory can be usefully delineated into strong and weak forms with regard to the causal mechanisms driving the observed correlations between typicality and valuation. These causal relationships are summarized in Figure 4-7.

The strong form of prototype theory emphasizes the primacy of prototype effects in valuation and corresponds to the version typically invoked by organizational theorists. Under this view, the ability of people to interpret and make sense of categorical combinations is a function of the (proto)typicality of those combinations. While prior work on prototype theory has not explicitly explained how prototypicality is linked to theories of value, a theory built on the primacy of prototype effects implies that people's theories of value are in largely determined by the prototypes they have available to them.[20]

[20] If we operate under the assumption that people are in fact confused by atypical offerings, it is interesting to consider whether people can construct or apply theories of value that do not use prototypes. After all, it is difficult to

However, as previously discussed, there are compelling reasons to doubt the claim that atypical offerings are confusing and that confusion necessarily entails a penalty. The weak form of prototype theory is an alternative interpretation that does not require these strong assumptions about confusion and the primacy of prototype effects. As shown on the right side of Figure 4-7, the weak version of prototype theory affords primacy to theories of value in determining valuation. Certain combinations become prototypical because they have been shown to perform well (Zuckerman 2017), thereby reversing the causal relationship between typicality and valuation assumed in the strong form. The weak form thus implies that typicality is more appropriately considered a correlate of value, rather than its direct and primary cause. Note that this view does not preclude audiences from being subject to prototype effects, but instead suggests that the constraints imposed by prototype effects on strategic choice and valuation are more limited than suggested by the strong form.

A particularly desirable property of the weak form of prototype theory is that it provides an explanation for the emergence and change of prototypes. By contrast, the strong version does not readily provide such an account. Taking the logic of the strong form to its extreme, some initial random conditions lead to some combinations becoming more typical, leading to a path dependent process where prototypes reproduce themselves and become entrenched.

As shown in the figure, both the strong and weak versions of prototype theory imply endogeneity issues that make causal identification challenging. A limitation of the empirical findings in this paper is that I am not able to address these concerns about causality. Of the two

---

see why someone would apply a theory of value that does not make sense to them. However, as Zuckerman [cite] argues, people have good reasons to apply theories of value that they do not believe in, or presumably even understand. Such scenarios can be expected when people care about third-order inferences (Correll et al. 2017). For example, in the context of the stock market, an investor may have a theory of value that is completely different from that of other investors. Yet, if prices are being driven by other investors' theories of value, then they would be wise to apply that theory.

forms, the weak version (shown on the right side of Figure 4-7) appears more straightforward to address in future research. In particular, longitudinal data could be used to test the temporal ordering of the causal relationship between coherence at time t=0 and typicality at t=1. The weak approach also appears more amenable to experimental tests, where valuation could be manipulated to see if it affects the subsequent typicality of combinations.

Typicality as an alternative type of coherence

The preceding discussion assumes that coherence and typicality effects are driven by two distinct processes, one rational and calculative, the other behavioral and automatic. This is also the approach taken by Boulogne and Durand (2021), who seek to synthesize prototype theory and the theory of value approach.

A second way of reconciling the relationship between coherence and typicality is to regard typicality as a distinct type of coherence. In this paper, theoretical coherence has referred to the extent to which a producer constructs an internally coherent identity over time. This type of coherence can be found in Wohl's (2019) study of artists, where she shows how they strive to construct consistent narratives around their evolving body of work. Whereas this kind of coherency can be regarded as coherency with respect to oneself, typicality may be regarded as coherency with respect to the field. Put differently, typical offerings are consistent with other established offerings and can readily be placed in a broader market context. Depending on an evaluator's goals, "self coherence" or "field coherence" may take primacy. For example, field coherence can represent salient focal points for coordination, particularly among strangers (Zuckerman 2017; Schelling). By contrast, *Twitch* viewers are often engaged in repeated interactions with specific streamers and their consumption is largely a private affair. Their primary concern is thus more likely to be about whether the streamer is consistent with the viewer's theory of value.

SUMMARY AND GENERAL DISCUSSION

Why are some combinations of categories more viable than others? Despite an extensive literature on categorization and valuation processes that has developed over the last several decades, surprisingly little guidance is to be found on this question. Prototype theory provides a clear standard for evaluating the viability of categorical combinations—that audiences penalize atypical combinations—but this explanation encounters significant theoretical and empirical difficulties. Extending the theory of value approach to categorization, I identify theoretical coherence—the extent to which a combination of categories appears coherent by someone applying a particular theory of value—as an alternative explanation of the viability of categorical combinations that overcomes the limitations of prototype theory. I develop and validate a novel empirical framework for identifying theories of value from community discourse and measuring the extent to which categorical combinations appear coherent when these theories are applied. Consistent with my theory, I find that audiences penalize incoherent combinations of categories and that these effects are robust to controlling for typicality. Importantly, I find that audiences impose greater penalties on offerings that appear incoherent from the perspective of their theory of value than offerings that appear incoherent from other closely related but distinct theories.

The analyses in this paper reveal the deeper underlying logics that structure relationships among categories, with important implications for our understanding of categorization and valuation processes. Key to this insight was exploiting a unique feature of my setting that afforded an unusually close comparison of two theories of skill being applied to the same set of offerings and in the same market context, allowing me to demonstrate how nuanced differences in otherwise similar theories of skill lead to consequentially different constraints on strategic choice and valuation. These findings are difficult to reconcile with prototype theory without imaginative ex post rationalization. Despite the theoretical and empirical challenges associated with prototype theory, it continues to undergird (often implicitly) contemporary research on categories and valuation. Moreover, although prototype theory and the theory of value strongly differ in their implications of how and why categories constrain strategic choice and valuation, these are often confounded. Beyond clarifying the challenges associated with prototype theory, one contribution of this paper is to provide a foundation for theoretically and empirically distinguishing these approaches.

That different logics can be used to guide categorical combinations also poses challenges to the notion of a generalist premium. Prior work in this area—within prototype theory as well as the theory of value approach—has almost exclusively focused on the conditions under which audiences are more or less tolerant of generalists and has largely ignored the question of which kinds of combinations are viable. A notable recent contribution in this regard is Goldberg, Hannan, and Kovacs (2016) complicate the notion of a generalist by pointing out an important distinction between tolerance for variety and tolerance for atypicality. A person who enjoys variety may dine at restaurants ranging from Ethiopian to Japanese yet demand that each restaurant be categorically pure. This kind of tolerance for variety stands in stark contrast to someone who seeks out Ethiopian-Japanese fusion cuisine. While their work is notable for problematizing the notion of generalists, the present work suggests that even Goldberg, Hannan, and Kovacs' (2016) distinction does not go far enough. The first-order question that producers and audiences grapple with is rarely one of "how much category spanning is acceptable," but "which combinations of categories are acceptable." The former question may take precedence in contexts where people apply a “logic of appropriateness” (March and Olson 1996) and value purity for its own sake. Yet such contexts are rarely the focus of categorization studies, which are often conducted in contexts where there is a strong “logic of consequences.” The crude distinction between specialists and generalists simultaneously over- and understates constraints on strategic choice and valuation. The staunchest purest is likely to accept some combinations, just as the most adventurous omnivore will not blindly consume any random combination. Rather than treating these as aberrations, they can be understood as the application of a particular theory of value.

Given that my concept of theoretical coherence was derived from the theory of value approach, it is no surprise that my findings are consistent with prior work on theories of value. Indeed, that different theories of value lead to different ways of organizing the category space is a basic insight of prior work on theories of value. For example, Paolella and Durand (2016) analyze the corporate law market and find that preferences for category-spanning law firms are contingent on the client’s theory of value: clients with more complex needs tend to view category-spanning law firms more favorably because they are deemed more capable of meeting their idiosyncratic needs.

However, the present work extends and clarifies the theory of value approach in several important respects. The first contribution to this literature is theoretical. The standard of theoretical coherence was largely implicit in prior work on theories of values. By developing and testing theoretical coherence, I make explicit and formalize one of the core insights of this approach. The second contribution is empirical: prior work has been limited in its ability to directly observe theories of value. Consequently, it has not been possible to make ex ante predictions about which categorical combinations are viable. By demonstrating how two theories of value applied to the same market context yield different valuations, I supply missing evidence for a core prediction of the theory of value approach.

The final contribution to the theory of value literature is methodological. This work provides an empirical framework for identifying theories of value and measuring theoretical coherence that can readily be applied to many settings. Although prior work has identified theories of value operating in their setting, it has been difficult to generalize these findings beyond the specific setting being studied. For example, in their analysis of the limits to diversification in corporate law, Phillips, Turco, and Zuckerman (2013) find that diversification into personal injury law triggers commitment concerns, whereas diversification into family law does not. This insight required deep familiarity with the setting and extensive qualitative data collection. But what might be equivalently problematic combinations in other settings? The empirical framework developed in this paper offers researchers tools to address this in their own work. The demands of this approach are modest, requiring only access to a large corpus of documents that reflects the theories of value of the communities of interest. As more archives are digitized and more community discourse moves online, such corpora are becoming increasingly available.

An important caveat of this computational approach to studying theories of value developed is that it does not remove the need for qualitative data and institutional knowledge. Word embedding models cannot be applied blindly and must be validated for their specific use case. Because the quality of word embeddings depends on the ends to which they are used, there is no one optimal word embedding. While some general-purpose tests do exist, such as analogical reasoning tasks, these are unlikely to be sufficient for the kinds of use cases most likely of interest to social scientists. For example, word embeddings that perform well on standardized analogical reasoning tests may perform poorly when used to study latent concepts like theories of value. Standardized tasks are useful for ensuring some baseline validity of the model, but ad hoc approaches to validation are likely needed in most social science applications. However, these ad hoc approaches likely require intimate knowledge of the setting being studied. Word embeddings can therefore not be regarded as a fully unsupervised approach, at least in the context of social science, but rather part of an iterative process of data collection and model tuning.

As word embedding models become more prevalent in social science research, additional work is needed to build better tools for validating word embeddings that address the needs of social scientists. The tanglegram used in this paper provides a simple graphical approach to validation, but does not constitute a formal test. One particular challenge that arises in designing ad hoc validation tasks is that the word embeddings may be validated against the outcome of interest, which would akin to p-hacking. Additional research on formal approaches to validating word embedding models would be particularly helpful in addressing such concerns. One notable development in this regard is the “Turing test” developed by Rodriguez and Spirling (2022), which compares the results of word embeddings to human judgments.

APPENDIX A: TANGLEGRAMS AS A SIMPLE BUT EFFECTIVE GRAPHICAL METHOD FOR COMPARING WORD EMBEDDING MODELS

With the relatively recent rise of word embedding models in the social sciences, few—if any—standardized methods have been developed for probing and comparing these models that cater specifically to the needs of social scientists who wish to use them to develop and test theory. I tried several other approaches before settling on the tanglegram as a simple but effective graphical method for comparing differences in theories of value across two communities. To my knowledge, this is the first use of tanglegrams for comparing the results across word embedding models. My hope is that this serves as a first step toward more powerful visual and formal methods for comparing word embedding models. To this end, I provide additional details about my implementation of the tanglegram and a brief comparison of the tanglegram to other common methods for probing word embedding models.

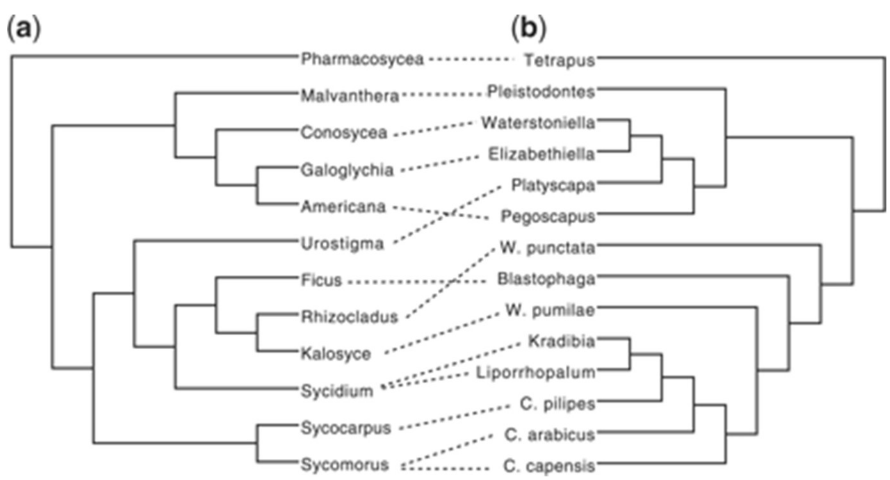

Figure A-1. Example of a typical tanglegram used to compare phylogenetic trees. The figure includes dendrograms. Source: https://bivi.co/visualisation/nn-tanglegram.

Tanglegrams are used to graphically evaluate the correspondence between two tree-like structures, such as those produced by hierarchical clustering algorithms, and are commonly used in biology to compare phylogenetic trees (Scornavacca, Zickmann, and Huson 2011). **Error!**

Reference source not found. shows a typical tanglegram used to compare phylogenetic trees. I depart from the traditional implementation in two ways. First, tanglegrams are traditionally ordered to minimize crossing lines (Matsen et al. 2016). This would not be appropriate in the present case because the rank ordering of the elements is meaningful and substantively important. Second, tanglegram labels are traditionally equidistant, with the distance between labels (and clusters of labels) shown by dendrograms (the tree structure). Since the present approach does not cluster the labels, this information cannot be communicated via a tree structure. Furthermore, the relative distance between labels readily conveys meaningful information about the relative importance of each term and is thus worth preserving.

One common approach to comparing the results of word embedding models is heatmaps. **Error! Reference source not found.** provides an example heatmap comparing the cosine distance between vector pairs. However, in the way heatmaps are used in this example, they cannot be used in the present case because distances between vector pairs in the two models is not of direct interest. Furthermore, vectors cannot be directly compared across models because they exist in different embedding spaces. While the latter issue could be overcome by aligning the embedding spaces, for example through Procrustes alignment (Hamilton, Leskovec, and Jurafsky 2018), this does not resolve the former issue. A more appropriate metric for the present purposes might be to take the difference in cosine similarity between the two models. However, showing these differences through shades of color is less clear than the vertical ordering and side-by-side comparison in the tanglegram.

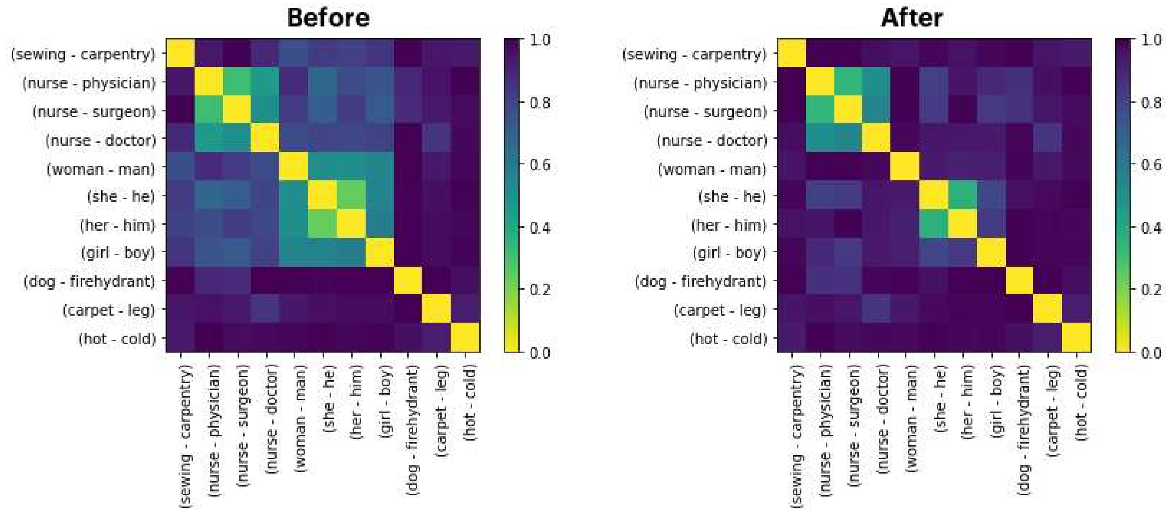

Figure A-2. Heatmaps comparing word embedding models before and after applying techniques for removing gender bias from embeddings. Values indicate cosine distance between vector pairs. Reproduced from *whatlies* https://koaning.github.io/whatlies/examples/lipstick-pig/.

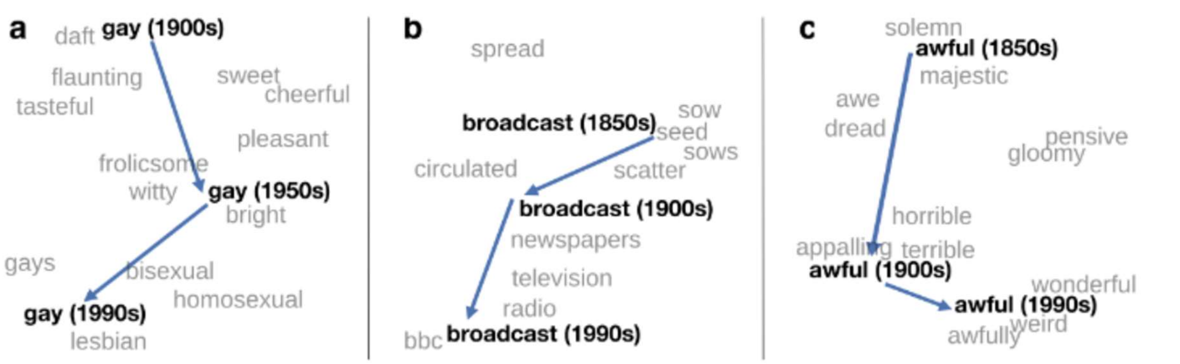

Figure 1: Two-dimensional visualization of semantic change in English using SGNS vectors.[2] **a**, The word *gay* shifted from meaning "cheerful" or "frolicsome" to referring to homosexuality. **b**, In the early 20th century *broadcast* referred to "casting out seeds"; with the rise of television and radio its meaning shifted to "transmitting signals". **c**, *Awful* underwent a process of pejoration, as it shifted from meaning "full of awe" to meaning "terrible or appalling" (Simpson et al., 1989).

Figure A-3. Showing movement in semantic space to compare differences in word sense across word embedding models. Reproduced from Hamilton, Leskovec, and Jurafsky (2018).

A second noteworthy approach is to show the movement of terms in semantic space, as illustrated in **Error! Reference source not found.**. However, this approach is more well suited to showing contextual differences in word sense across models. In the present case, any differences in word sense are minimal across the two corpora. “Aim” and “mechanics” mean the same thing in both communities.

Finally, **Error! Reference source not found.** presents the top 25 terms most closely associated with skill in eSports and speedrunning as a scatterplot. This method is capable of conveying the relationship between terms in the models, unlike the heatmap and semantic change plots. However, this approach is less appropriate than the tanglegram for two reasons. First, the vertical orientation of the tanglegram conveys the rank ordering of terms more clearly. Second, this plot implies that terms should be evaluated off the 45 degree line. However, it is not clear what the slope of this line should be. In the case that the rank ordering of terms is consistent across models but shifted by a constant, evaluating terms off the 45 degree line would misrepresent the correspondance between the two models.

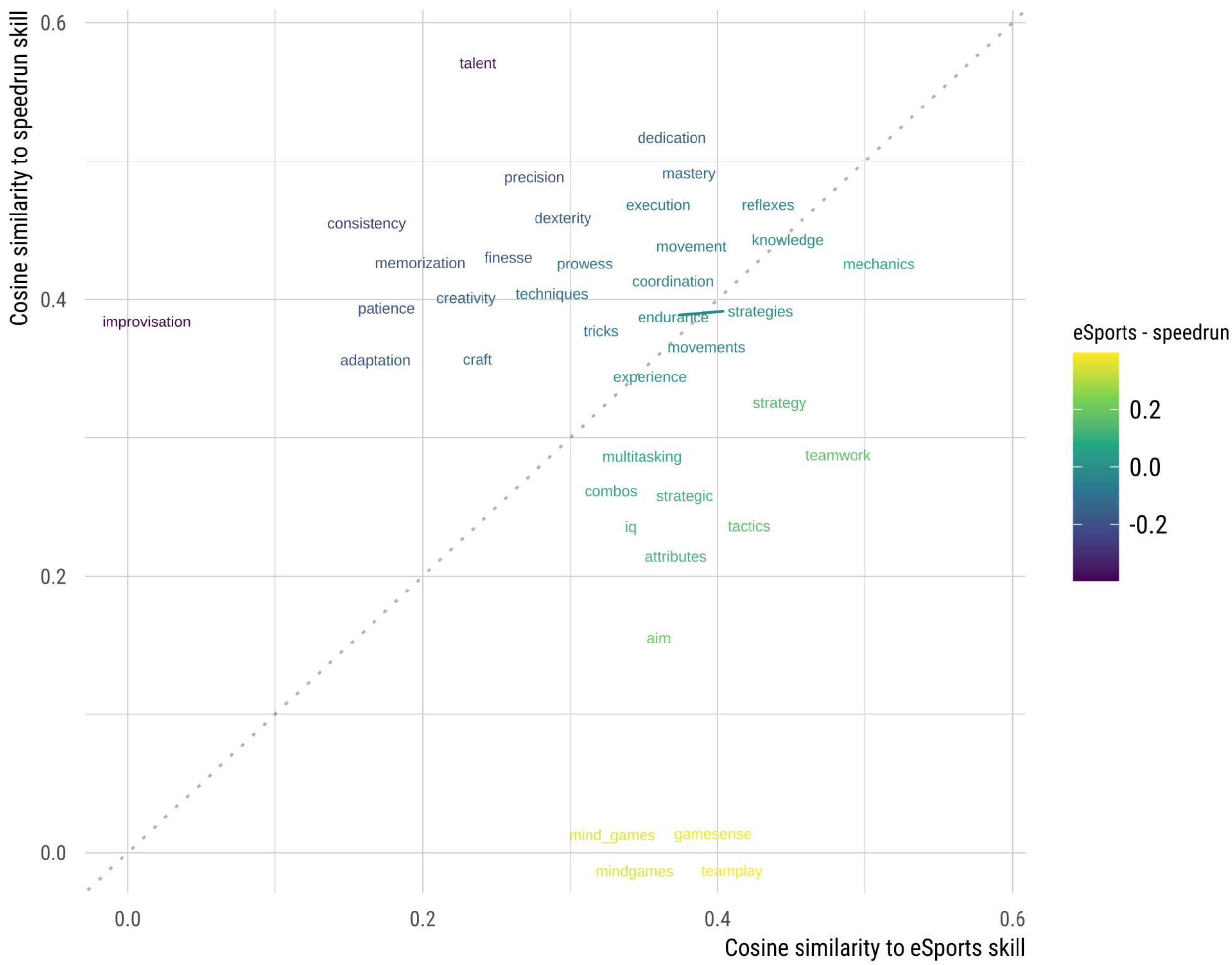

Figure A-4. Scatterplot comparing results of word embedding models trained on eSports and speedrun corpora. The figure shows the top 25 terms most closely associated with the vector "skill + skills," measured by cosine similarity.

APPENDIX B: REDDIT CORPORA EXCLUSION CRITERIA

Table B-1. Changes to full *Reddit* corpus after applying exclusion criteria.

Sample	Submissions	Comments	Total posts	% of full sample
Full	169,471,768	11,845,536	181,317,304	100.0%
Exclude moderator posts	168,427,647	11,831,257	180,258,904	99.4%
Exclude posts < 10 words	113,584,783	6,572,778	120,157,561	66.3%
Exclude repetitive content	112,687,332	6,410,090	119,097,422	65.7%

APPENDIX C: SAMPLE CONSTRUCTION DETAILS

Identifying eSports and speedrun streamers

I identified eSports and speedrun streamers by extracting links to *Twitch* channels from player profiles on *Liquipedia* and *Speedrun.com.* Player profiles on these sites contain links to the player's social media accounts, including their *Twitch* channel if one exists. Streamers who have profiles on these sites have an incentive to ensure their profiles contain accurate links to their social media accounts because it provides a way for viewers to discover their channels. I thus used these links to identify eSports and speedrun streamers.

An alternative approach to identifying eSports and speedrun streamers is to use the tags they apply to their streams, such as "Competitive" and "Speedrun." However, there are two disadvantages to this approach. First, in the case of eSports, the number of tags is limited and applied inconsistently by streamers. Second, this approach requires researcher judgment in categorizing streamers. My present approach overcomes these limitations.

In total, I identified 12,178 eSports streamers, 22,240 speedrun streamers, and 62 streamers who appeared in both datasets. The streamers identified through these sources thus involve self-selection into eSports and speedrunning activity as recognition by the communities that their participation in those activities was legitimate. It is interesting to note one subtle difference in the categorization processes: whereas the eSports streamers primarily reflect community categorization, the speedrun streamers primarily reflect self-announcements. A producer who announces themselves as being of a certain type ostensibly seeks to be recognized and evaluated according to that type. By contrast, producers who are categorized by others may wish to be recognized as belonging to a different category. This tension between a producer's desired identity and the identity ascribed to them by their audience is at the heart typecasting (Zuckerman et al. 2003).

Liquipedia data

Liquipedia is a *Wikipedia*-style encyclopedia of eSports. The site is maintained by a community of eSports enthusiasts and documents eSports tournaments and associated people at both the amateur and professional levels across 44 games. Although there are other sites dedicated to eSports statistics in specific games, *Liquipedia* is particularly useful because of the breadth of games it covers as well as its size (3.4 million users and 12.3 million edits on 1.4 million pages as of March 2022). I excluded profiles of people who did not have a history of competing in eSports, such as team owners and commentators. Player profiles thus represent people who have competed in an eSports competition deemed by the community to meet a minimum level of noteworthiness.

Table C-1. eSports streamers identified on *Liquipedia*.

Game	Total player profiles	Profiles with Twitch channel	Prop. with stream
StarCraft2	5360	834	0.16
Rocket League	3668	1365	0.37
Counter-Strike	3466	1432	0.41
Smash	2566	906	0.35
League of Legends	2382	897	0.38
PUBG	2260	616	0.27
Overwatch	2080	1077	0.52
Fighters	1711	0	0
Valorant	1624	1079	0.66
Apex Legends	1264	939	0.74
Hearthstone	1205	423	0.35
Pokemon	1151	98	0.09
Dota 2	1141	412	0.36
Warcraft	1073	284	0.26
Rainbow Six	1042	536	0.51
Heroes of the Storm	823	304	0.37
Arena FPS	627	152	0.24
StarCraft	595	120	0.2
Age of Empires	571	268	0.47
Fortnite	560	396	0.71
Call of Duty	535	180	0.34
Arena of Valor	456	15	0.03
Clash Royale	442	64	0.14
Trackmania	411	126	0.31
Free Fire	402	25	0.06
League of Legends: Wild Rift	336	69	0.21
Team Fortress	325	165	0.51
Halo	287	0	0
mobilelegends	259	0	0
FIFA	233	140	0.6

World of Warcraft	230	106	0.46
Artifact	139	41	0.29
Battalion	137	51	0.37
Runeterra	126	73	0.58
Sim Racing	104	5	0.05
Teamfight Tactics	94	82	0.87
Brawlstars	83	16	0.19
Battlerite	68	14	0.21
Paladins	46	39	0.85
Star Wars: Squadrons	43	0	0
Dota Underlords	40	22	0.55
Magic: The Gathering	39	31	0.79
Crossfire	36	6	0.17
splitgate	7	6	0.86
Auto Chess	3	0	0
sideswipe	1	0	0
Total (non-unique)	40051	13414	0.33

Speedrun.com data

Speedrun.com is the de facto leaderboard for speedrunning attempts. A player who attempts a speedrun can submit a recording of their attempt to the site. Designated moderators for that game then verify the integrity of the run. Users do not have to submit a new record to be included in the data, but can submit any attempt. As of March 2022, the site is comprised of 2.8 million runs by 1.1 million users in 27.6 thousand games. Whereas the bar for inclusion in *Liquipedia* is competing in a recognized tournament, the bar for inclusion in the *Speedrun.com* data is merely submitting a single run. Because users with only a few submissions are unlikely to be considered speedrunners, I limited my sample to users who had submitted at least 10 non-rejected runs (thus including runs pending verification).

The *speedrun.com* data consist of all users with valid Twitch streams who submitted at least 10 valid runs. I applied several exclusion criteria to construct the sample. Table C-2 shows descriptive statistics of the sample after applying each exclusion criterion. First, I removed all runs that were rejected. This was done to ensure that only legitimate runs are considered. Most runs are rejected for being duplicates, missing video proof, or for being spam. Runs can also be rejected when they do not meet the specific guidelines for the category the run is submitted to. Next, I removed runs from players who had submitted less than 10 runs as these players are unlikely to engage in speedrunning on a regular basis. Finally, I limited the sample to players who had valid links to Twitch profiles. I also removed players who had valid Twitch profiles but where the same Twitch profile was associated with multiple Speedrun.com user accounts. These were primarily junk responses, such as 'nope' or 'none.'[21] The final sample consists of 22,302 Twitch streamers who have submitted 1,202,977 runs across 22,096 games.

Table C-2. Changes to *speedrun.com* sample after applying exclusion criteria.

Sample	**Submitted runs**	**Unique players**	**Unique games**
All runs	2,611,051	314,387	26,552
Exclude rejected runs	2,352,635	293,042	26,150
Exclude players with < 10 submissions	1,791,574	39,722	24,053
Exclude players without Twitch streams	1,202,977	22,302	22,096

[21] The following Twitch usernames were excluded: 'nope', 'none', 'open', 'roku', 'Roblox', 'TWiT', 'twith', 'Twitch_TV', 'GitHub', 'twitvh', 'twtch', 'i_dont_have_twitch', 'Twitxh', 'twicth', 'twitchtvcom', 'Twirch', 'switch', 'twtichtv', 'Twitct', 'htpps', 'LinkedIN', 'i_dont_have'. I also removed 7 users who had multiple Speedrun.com accounts.

CHAPTER 5

AUDIENCE COMMITMENT AND THE DYNAMICS OF ENTREPRENEURIAL EXPERIMENTATION

INTRODUCTION

The central challenge facing the entrepreneur is that the viability of their idea is inherently unknowable in advance (Hayek 1968; Kirzner 1997). Indeed, if the viability of an idea could readily be determined, entrepreneurs would face little risk but also little opportunity.

Entrepreneurial experiments are traditionally thought to improve outcomes because they avoid commitment and create real options (Manso 2016; Nanda and Rhodes-Kopf 2016; Ries 2011). Rather than make costly commitments to developing a product for which there may be no demand, entrepreneurs can learn about the value of their idea and iteratively refine it through a series of low-cost experiments. Experiments are thought to improve outcomes in two ways. First, experiments provide an early signal of the viability of the idea, giving the entrepreneur the option to pivot (including exit) or continue development (Manso 2016; Nanda and Rhodes-Kopf 2016; Trigeorgis and Reur 2017). Second, experimentation yields learning about the environment, resulting in the identification of more attractive markets and better product-market fit (Murray and Tripsas 2014).

Nonetheless, an important recent development in the literature on entrepreneurship is that experiments are not simply low-cost tests, but in fact create partial commitments to a course of action (Gans, Stern, and Wu 2019; Chavda 2018). The key insight of this work is that experimentation is not a panacea for reducing uncertainty but involves tradeoffs. For example, Chavda (2018) demonstrates this point empirically in the context of television shows. He finds that television shows that use pilot episodes (a low-cost way to gauge viability) attract less famous actors relative to shows commissioned for an entire season. While the use of less famous actors reduces the costs of producing a pilot, it threatens the long-term success of the show if it goes into production because shows with less famous actors do not attract as many viewers as shows with more famous actors.

In this paper, I complement work on the "experimentation entails commitment" view by documenting a situation where audiences demand commitment before they will tolerate experimentation. I hypothesize that commitment to an audience presents a dilemma for entrepreneurs. On the one hand, audiences are more likely to tolerate experimentation and provide feedback when the entrepreneur is committed to them. On the other hand, obligations to heed this feedback reinforce the entrepreneur's commitment to the audience, making it more difficult to diversify into new markets. One market context in which these dynamics are likely to occur is when identity is a salient factor shaping demand, such as in cultural markets. This tradeoff views experimentation as a relational process and highlights the role of demand-side factors in shaping the dynamics of entrepreneurial experimentation. A key insight is that experimentation is costly for participants and that recruiting tolerant participants can represent a significant barrier to experimentation, especially for new, resource-constrained entrepreneurs.

The rest of the paper is organized as follows. In the next section, I introduce the setting for the present study, *Twitch*—a leading livestreaming platform where cultural entrepreneurs regularly experiment with new additions to their repertoires. Drawing on a case study of one entrepreneur's experience experimenting on the platform, I develop theory to identify and clarify the tradeoff between audience commitment and growth. I then provide a preliminary test of key aspects of the theory using quantitative data from my setting.

SETTING

As described below, streamers are a type of cultural entrepreneur who engage in experimentation. Because of the highly interactive and social nature of Twitch, it is well-suited to the present study because it shines a light on the role of audience interaction in experimentation.

Competing for scarce attention means that streamers face continuous pressure to update their repertoire. In the case of *Twitch*, streamers' repertoires primarily consist of a variety of video games. Constructing a repertoire of games is challenging, however, because introducing a new game often results in viewer drop off. Streamers must therefore be strategic about which games they add and how they add them. To this end, many streamers engage in experimentation. For example, SimCopter1, a professional streamer, offers the following advice to new streamers seeking to grow:

> Don't be afraid to experiment. Not every game you play will be a home run, it can take a while to find such a game, having a rotation of three to six games helps keep things fresh while creating a routine for you and your audience. Kick out an underperforming game every once in a while and bring in a new one. It may be your next home run. (Twitch Creator Camp 2018).

In addition to these substantively important features of the context, the *Twitch* data used in the present study provide several empirical benefits that allow an unusually close look at the dynamics of entrepreneurial experimentation. One challenge with studying entrepreneurial experimentation is that it many entrepreneurs and their experiments are invisible to researchers. The data used in the present study include observations on all streamers over an extended period. Furthermore, because the data consist of fine-grained longitudinal data on individual streamers, I can observe even very brief instances of experimentation.

THEORY

To develop and contextualize the theory, I introduce the case of SimCopter1,[22] a professional *Twitch* streamer. This case was chosen for its richness in describing one streamer's experience experimenting on *Twitch*. In using this case to develop theory, I do not claim that the case is representative of all streamers. Instead, it should be interpreted more narrowly as support for the existence of the dynamics under consideration. Whether these dynamics do in fact hold in the broader population will be tested using data on all streamers.

To situate the case of SimCopter1, recall that the focus of this study is an entrepreneur who wishes to conduct a small-scale test of an idea, receive feedback on it, and then decide whether to continue to pursue the idea or pivot to an alternative course of action (including exit). The entrepreneur wishes to conduct the test at the lowest cost that will permit a useful signal of its viability—often referred to as a "minimum viable product" or MVP (e.g. Ries 2011).

However, simply launching an MVP may not result in feedback that is useful for making the decision to pivot or persevere. SimCopter1's initial attempts at experimentation illustrate this point:

[22] The quotations from this case were drawn from a video from the Twitch Creator Camp (2018) in which SimCopter1 is providing advice to new streamers on how to grow their channels. The video is from Twitch's "Creator Camp," which is a free resource to help new streamers establish themselves by adopting best practices and learning from the experience of more established streamers. SimCopter1 began streaming on Twitch in 2014. In July 2018 (the month prior to making the Creator Camp video) he had 46,100 followers and 394 average viewers per stream.

> After a while I felt there may be greener pastures, so I experimented with different games. But I didn't understand that it took time to adjust your audience to other games, and I got pretty disheartened at times…when I when I first tried other games… they fell flat… I simply hoped they would watch. Most didn't and it hurt. (Twitch Creator Camp 2018)

Following these initial failed attempts at experimentation, SimCopter1 returned to the game he started with, *Hearthstone*. He exclusively streamed this game for several more years. It wasn't until an audience member provided feedback to him that allowed him to understand why his earlier experiments had failed:

> I remember at Pax South 2016 [a large gaming convention], I spoke with a [Twitch] staff member and I told him what I felt was going right and wrong with Hearthstone in my broadcast and he told me--this paraphrased, by the way--"I feel like you're at your best with single-player games where you can pause and take time to explain things. It feels like we're all playing the same game together." And this is all kind of like in the scope of me playing a lot of Hearthstone, which is a multiplayer, dueling competitive game, right? He's telling me my strengths were something very different than what I was showing most of the time. So, it was a bit of a shock after that, I kind of realized the core strength about me that I couldn't see…it took others to help me realize what I was best at. (Twitch Creator Camp 2018)

An experiment could have succeeded or failed for any number of reasons that may not be obvious to the experimenter. Consequently, customer feedback is essential to refining the MVP and understanding *why* it succeeded or failed and what should be done next. The ideal feedback is rich and timely, allowing for a rapid iteration cycle (Ries 2011). If customers continue to respond positively to the MVP, it can be incrementally scaled up until the entrepreneur decides to fully commit to the offering. The importance of audience feedback is reflected in the advice that SimCopter1 gives a new streamer:

> … reach out to your community and see, you know, ask them very literally, "what do you guys think of that game? Did you like it?" They might say, "yeah, awesome game, but I don't know if the way you're broadcasting it really caught me," or maybe vice versa. "We love you, but this game is ehh, it's OK." (Twitch Creator Camp 2018)

An underrecognized issue in prior work on entrepreneurial experimentation is that experiments impose costs on participants. The entrepreneur seeks to conduct tests using MVPs, which are typically crude prototypes that contain errors and limited functionality. Participants must not only endure crude prototypes, but also invest time and resources in providing usable, timely feedback. Consequently, users are unlikely to be willing to participate in these experiments if they do not believe the product will eventually reach a stage where it will be useful to them. This situation requires that the participants are confident in the entrepreneur's capability in developing the product (this includes access to necessary financial, physical, and human capital) as well as the entrepreneur's commitment to the participants.

SimCopter1 describes how audiences vary in their tolerance for experimentation. He notes that attempts at experimentation simply result in viewer exit if the streamer does not invest the time in preparing[23] the audience and when the audience lacks a prerequisite level of trust in the streamer (emphasis added):

> It takes a few times for people to kind of understand and know [the new game]. But basically, the more you [introduce new games], the easier it gets. The more you do it, the more smoothly you do it. And the more clearly you indicate to your audience that [the new game] is something that will interest them or could interest them, the more likely they are to be like, **"all right, I'll stick around. I'll stick around. OK, you got my trust,"** you know. But if you're new and you're smaller and you're not quite established, the hit might be taken harder. But don't take it personally. It's not. It's just a part of being a smaller and newer broadcaster. People are not going to be as willing to stick around, but you gain momentum. It snowballs over time... There's no silver bullet when it comes to streaming. (Twitch Creator Camp 2018, emphasis added)

Reinforcing the notion that audience commitment creates an environment that is supportive of experimentation, he notes “whenever you're switching games and switching content is going to be a bit of a viewer drop off and whatnot, but the people who love you will stick around.” (Twitch Creator Camp 2018).

It is useful to distinguish between two kinds of commitment. Literature on strategy and entrepreneurial experimentation considers commitment broadly as an investment that is difficult or costly to reverse (e.g., Ghemawat 1991). For example, a firm that invests in excess capacity can deter entry by other firms because it has made a credible commitment to increasing supply

[23] The literature on entrepreneurial experimentation has largely focused on the importance of *receiving* customer feedback. An interesting aspect of this case is that it highlights the importance of communicating *to* customers. In sum, these complementary processes highlight the relational nature of experimentation.

(Dixit 1980). In the context of experimentation, creating a pilot version of a television show entails a commitment to use the same actors for the entire season if the show launches (Chavda 2018).

By contrast, work by economic and organizational sociologists considers commitment more narrowly and focuses on relational commitments (e.g., Phillips et al. 2013; Hahl and Ha 2020)[24]. In these cases, audiences demand commitment from producers because they are engaged in a principal-agent relationship. Intending to serve another audience will be interpreted as an act of betrayal and penalized, even when the producer's capability for serving the original audience is not affected. Such demand for commitment is frequently observed in identity-heavy industries. For example, audiences penalize producers of craft beer when they are acquired by mass market producers or when seeking to serve the mass market (Carroll and Swaminathan 2000; Frake 2017).

Returning to the context of experimentation, audiences are less likely to tolerate experimentation from producers who they believe to be uncommitted to them, even if they believe the entrepreneur is capable of ultimately delivering the desired product. This unwillingness is because the effort expended by the audience in enduring crude prototypes and providing feedback will be foregone if the entrepreneur decides to pivot to a new market segment. Indeed, we might expect audiences to be generally wary of early-stage experimentation because the costs of participation are higher (due to cruder prototypes) and the purpose of such experimentation is often to explore the value of different market segments (Ries 2011). In contexts like *Twitch*, where relationships between producer and audience are especially salient

[24] I would like to thank Minjae Kim for sharing work in progress on this topic.

(Blight 2018), producers who ignore feedback or use it to pivot to another market are particularly unpalatable because doing so is interpreted as an act of betrayal (Phillips et al. 2013).

To summarize the previous discussion, the challenge for the entrepreneur is that experiments cannot be conducted unilaterally but require tolerant participants. Because experiments impose costs on participants, they will vary in their tolerance for experimentation according to their perceptions of those costs. Demonstrating capability and commitment becomes increasingly important for securing tolerant participants when prototypes are crude, more feedback is required, and when audiences interpret a pivot to a new audience as an act of betrayal.

The tradeoff between strong and weak audience commitment strategies

Having explicated the need to commit to an audience before they will tolerate experimentation, we now turn to understanding the tradeoff this situation poses for entrepreneurial experimentation. Let us consider two idealized commitment strategies: a strong audience-commitment strategy and a weak audience-commitment strategy (Anjos and Reagans 2013).[25] Entrepreneurs who follow a weak commitment strategy invest few resources in building a relationship with an audience and continuously seek out more valuable audiences; entrepreneurs who follow a strong commitment strategy invest significant resources in building a relationship with a specific audience and seek to develop valuable opportunities within the context of that audience.

[25] Anjos and Reagans (2013) consider a mixed commitment strategy whereby the producer enacts strong commitment to some ties and weak commitment to other ties. In this paper, I consider a context where the producer (entrepreneur) faces an undifferentiated (inseparable) audience and thus cannot adopt a mixed strategy. The strategic implications of separable vs. inseparable audiences are discussed in Zuckerman et al. (2003).

In the first instance, consider an entrepreneur who follows a strong commitment strategy. Insofar as this entrepreneur credibly signals commitment to their audience, they face participants who are more tolerant of experimentation (a similar logic is at the heart of much work on identity-based limits to diversification, e.g Phillips et al. 2013; Hahl and Ha 2020). Consequently, the entrepreneur can conduct a greater number of experiments and receive feedback that is richer and timelier, thereby facilitating faster iteration cycles (Uzzi and Lancaster 2003; Reagans and McEvily 2003). Note that because we are considering a situation where the audience is only willing to endure experimentation and provide feedback because they believe the producer is committed to them, it follows that the audience expects the producer to heed the feedback. Ignoring the feedback, or worse yet, using it to serve another market, will be interpreted as an act of betrayal. Insofar as feedback is intended to serve the participants (rather than a different audience) and the strong commitment strategy entails an obligation to heed the feedback, the entrepreneur is led down a path that reinforces product-market fit with their current audience. To the extent that participants systematically differ from other markets, experimentation conducted in the context of a strong commitment strategy shapes the dynamics of experimentation in a way that creates value on the intensive margin (i.e., for the existing audience) and less so on the extensive margin (i.e., for new audiences).

By contrast, consider an entrepreneur who follows a weak commitment strategy. While this entrepreneur does not have the same obligation to heed feedback, she will find it more difficult to attract audiences who are tolerant of experimentation. Compared to an entrepreneur following a strong commitment strategy, however, learning that does take place can be more readily applied to entering new markets—i.e., on the extensive margin. In short, a weak commitment strategy is expected to result in experimental dynamics that prioritize value creation on the extensive margin, whereas a strong commitment strategy is expected to result in experimental dynamics that prioritize value creation on the intensive margin.

One interesting implication of this tradeoff relates to the "cold start" problem. If entrepreneurs must experiment to identify valuable ideas, how do they start testing their ideas when no one is willing to provide sufficient feedback? The tradeoff suggests that the entrepreneur can achieve a given level of feedback by either committing to an audience so that they will accept a cruder prototype, or by committing to refining the prototype to a stage where the audience is willing to test it and provide feedback.

EMPIRICAL STRATEGY

This paper tests two implications of the theory that audience-commitment involves a tradeoff for experimentation. First, I test whether streamers who follow a strong audience-commitment strategy have viewers who are more tolerant of experimentation. Second, I test whether streamers who follow a strong audience-commitment strategy conduct experiments that lead to lower growth on the extensive margin.

It is difficult to measure the extent to which an audience believes a streamer is committed to them. Recall that credible commitments entail investments that are difficult or costly to reverse. Thus, I operationalize commitment as the extent to which a streamer makes sunk-cost investments into their audience. In markets like *Twitch* where producer-audience relationships are salient, relational labor (intimate work to connect with one's audience) is a prominent form of these kinds of irreversible investments (Baym 2018l; Taylor 2018).

To this end, I exploit a unique feature of my setting that allows me to observe relational labor: time spent streaming in the "Just Chatting" category. As the name implies, Just Chatting streams focus on streamer-viewer interaction and not on a particular activity, such as a game.

Just Chatting has grown to become the most popular category on *Twitch*. While some streamers exclusively stream in Just Chatting, other streamers will intersperse Just Chatting throughout their streams. For example, a common pattern is to begin a streaming session in Just Chatting before transitioning to a game.

Hypotheses

Before proceeding with testing my theory, I test a key assumption of the theory: that experimentation is costly for participants and that participants display at least some intolerance. I test this assumption by looking at whether entering a new category results in viewer drop-off. More specifically, I test whether the first 15-minute slice when a streamer plays a game has lower viewership than other slices. The more time a streamer spends in a category, the greater the opportunity to be discovered by viewers of that category. Thus, by limiting attention to this first slice, I am better able to observe the effect of experimentation on the intensive margin and avoid capturing any growth that may occur on the extensive margin:

> ***H1 (intolerance for experimentation):*** *Experimenting with a new category results in viewer-drop off.*

Next, I test whether a strong audience-commitment strategy increases tolerance for experimentation. More specifically, I test whether relational labor moderates the relationship between experimentation and viewer drop-off. Insofar as a strong audience-commitment strategy increases tolerance for experimentation, we should expect streamers who engage in relational labor immediately prior to experimentation to experience less drop-off:

H2 (commitment increases tolerance for experimentation): *The viewer drop-off associated with experimenting with a new category is less when the streamer engages in relational labor immediately prior to experimentation.*

Finally, I test whether experimentation conducted by streamers following a strong audience-commitment strategy reduces acquisition of new viewers. Entering a new category represents an opportunity for the streamer to be discovered by viewers of that category. A strong commitment strategy can limit the ability of a streamer to attract viewers in a new category because it results in a playstyle or streaming style that is tailored to existing viewers rather than new viewers.

H3a (extensive margin, viewer count): *The more relational labor a streamer engages in, the fewer viewers they acquire when entering a new category.*

As a robustness check, I test whether strong audience-commitment strategies result in fewer additional followers from entering new categories.

H3b (extensive margin, followers): *The more relational labor a streamer engages in, the fewer followers they acquire when entering a new category.*

DATA

Data collection and sample construction

The data consists of all live streams that occurred on the platform between November 2020 and February 2021.

To make initial analyses more tractable, I limit my sample in two ways. First, I limit my sample to the cohort of streamers who joined *Twitch* during the observation period. An added benefit of this subset is that the sample contains only streamers for whom I have the entire history. Second, I limit my sample to streamers who are likely to be strategic about experimentation and who face potential rewards and consequences for experimentation. many streamers who joined Twitch during the observation period only try streaming a few times and are not strategic about their experimentation. Furthermore, many new streamers have an average viewership between 0 and 1. Thus, experimentation for these streamers is largely inconsequential as there is virtually no opportunity for variation in the dependent variable. I therefore limit my sample sample to streamers who achieved "affiliate" status during the observation period. *Twitch* awards streamers with "affiliate" status after reaching a modest performance threshold.[26] I do this because

Summary statistics

Summary statistics are presented in Table 5-1 and correlations are presented in Table **5-2**. The final sample consists of 15,290,838 observations of 42,330 unique streamers; 1,761,730 streams featuring; and 7,938 unique categories (the vast majority of which are distinct games). The average streamer in this sample streamed for a total of 90.3 hours, held 41.6 streaming sessions, and played 6.0 unique games. Furthermore, the average streamer had an average of 17.8 concurrent viewers and 283.7 followers.

[26] The requirements for affiliate status are "at least 500 total minutes broadcast in the last 30 days, at least 7 unique broadcast days in the last 30 days, an average of 3 concurrent viewers or more over the last 30 days, and at least 50 followers" (*Twitch* website). One objection to this approach is that restricting the sample to streamers who reach affiliate status results in selection on the dependent variable. However, two substantive considerations ameliorate these concerns to some extent. First, affiliate status represents a "minimum" performance threshold at which we may begin to consider the behavior of streamers to be relevant for the present paper. Many streamers excluded by this criterion are not engaged in strategic behavior aimed at building an audience. In particular, as is common on digital platforms, many who join the platform to simply try streaming once or twice, never to return. Second, even among the streamers who fall below this performance threshold but are in fact strategic in their attempts to build an audience, including them presents a challenge for modeling viewer drop-off because the modal number of viewers they are at risk of losing is 0. Future analyses will explicitly model this limited dependent variable.

Table 5-1. Summary statistics.

	Mean	**Std. Dev.**	**Min**	**Max**
Viewer count (log)	1.65	0.95	0.00	11.43
Follower count (log)	4.33	1.46	0.00	12.79
First session in new category	0.01	0.11	0.00	1.00
Prior session was "Just Chatting"	0.08	0.27	0.00	1.00
Unique games in repertoire (log)	1.55	0.70	0.69	4.50
Cumulative hours in "Just Chatting" (log)	0.68	1.28	0.00	7.28
Cumulative hours streamed (log)	4.07	1.24	0.22	7.73
Category viewers (log)	9.93	2.51	0.00	14.84

Table 5-2. Correlation table.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
1. Viewer count (log)	1						
2. Follower count (log)	0.49	1					
3. First session in new category	-0.04	-0.05	1				
4. Prior session was "Just Chatting"	0.20	0.13	0.01	1			
5. Unique games in repertoire (log)	0.03	0.17	0.06	-0.02	1		
6. Cumulative hours in "Just Chatting" (log)	0.27	0.28	-0.01	0.62	0.26	1	
7. Cumulative hours streamed (log)	0.14	0.54	-0.07	0.03	0.35	0.27	1
8. Category viewers (log)	0.03	0.07	-0.09	0.30	-0.18	0.17	-0.01

Measures

The outcome variables are *viewer count (log)* and *follower count (log)*. Viewer count refers to the number of concurrent viewers at the time the stream was observed. *Followers* refers to the number of viewers who follow the channel. Following a channel is free and is primarily used to get a notification when the streamer goes online. This contrasts with a *subscription,* for which the viewer pays a monthly fee. Both variables are log transformed.

The primary measure of audience commitment is time spent in the “Just Chatting” category. As discussed previously, a primary way in which streamers generate commitment is by spending time in the “Just Chatting” category. Time spent in Just Chatting represents relational labor, which is an audience-specific sunk cost and thereby a measure of commitment. While streamers can and do engage in relational labor while in other categories (e.g., answering audience questions while playing a game), Just Chatting represents time solely dedicated to this activity. Thus, on average, streams in Just Chatting are more likely to be engaging in more intensive relational labor than streamers in other categories.

Cumulative hours in "Just Chatting" (log) is the total number of hours (logged) that the streamer has spent in Just Chatting. Streamers are [assumed] to appear more committed to their audiences when they spend more time in Just Chatting.

Prior session was "Just Chatting" is a dummy variable that takes one when the streamer was in the “Just Chatting” category in the prior observation. This variable indicates that the streamer was engaged in some form of relational labor prior to introducing a new game by explaining to the audience how it fit and was more likely to be preparing the audience for the transition to the new game compared to a streamer who didn’t engage in “Just Chatting.”

First session in new category is a dummy variable that takes one for the first observation in which a streamer enters a new category they have never previously streamed in. In this case, “new” means that the streamer has never played the game before, not that the game itself was recently released. Note that the variable takes one only for the first “slice,” such that it captures viewer count in the first 15 minutes of playing a new game. As a result, any change in viewer count is likely to be driven by viewers exiting the stream. If the variable captured subsequent slices, the streamer would be discoverable in that new category for a longer period, thereby potentially leading to an influx of viewers looking for streamers in the category.

Unique games in repertoire (log) captures the number of games in the streamer's repertoire. Each game can be considered a standalone experiment. The streamer can introduce a game, observe its performance, gather audience feedback, and then decide to pivot or persevere.

I also include several time-varying controls that affect viewers and followers. *Cumulative hours streamed (log)* is the total number of hours streamed. Viewer and follower count is strongly associated with time spent on the platform. *Category viewers (log)* is the total number of viewers of all streamers in the present category. Larger categories have greater potential to attract new viewers.

RESULTS

Table 5-3 presents fixed-effects regression results. All models include streamer (individual) fixed effects. The results support the hypotheses.

Models 1 and 2 test the effect of experimenting with a new game on viewer count. Model 1 is a baseline model; model 2 adds control variables. Consistent with Hypothesis 1, I find that the first session associated with a new game is associated with a decrease in viewer count. Consistent with Hypothesis 2, I find that engaging in relational labor immediately prior to introducing a new game results in fewer viewers exiting.

Models 3 and 4 test the moderating effect of audience commitment (i.e., relational labor) on the relationship between expanding one's repertoire of games and viewer count. Consistent with Hypothesis 3a, I find that streamers who enact a stronger audience commitment strategy attract fewer new viewers for each new market (game) they enter compared to streamers who enact a weak commitment strategy.

Table 5-3. Effect of high audience-commitment strategy on viewership and followers: Fixed effects analysis.

	Outcome: Viewer count (log)				*Outcome: Follower count (log)*	
	(1)	(2)	(3)	(4)	(5)	(6)
First session in a new category	-0.336***	-0.276***				
Prior session was "Just Chatting"	0.113***	0.108***				
First session in a new category * Prior session was "Just Chatting"	0.048***	0.015***				
Unique games in repertoire (log)		-0.008***	0.284***	-0.021***	1.778***	0.164***
Cum. hours in "Just Chatting" (log)			0.243***	0.106***	0.858***	0.119***
Unique games in repertoire (log) * Cum. hours in "Just Chatting"			-0.066***	-0.021***	-0.281***	-0.040***
Cum. hours streamed (log)		0.170***		0.160***		0.848***
Category viewers (log)		-0.006***		-0.003***		0.001***
Constant	1.645***	1.030***	1.129***	1.014***	1.351***	0.589***
Individual fixed effect	Yes	Yes	Yes	Yes	Yes	Yes
N	15,290,838	15,290,838	15,290,838	15,290,838	15,290,838	15,290,838
R2	0.005	0.064	0.041	0.063	0.455	0.707
Adj. R2	0.002	0.061	0.039	0.060	0.454	0.706

Models 5 and 6 differ from 3 and 4 only in that the dependent variable is the number of followers the streamer has. These tests are included as a robustness check. Consistent with Hypothesis 3b, I find that streamers who enact a strong commitment strategy acquire fewer followers per game added to their repertoire.

Overall, the preliminary results presented here lend validity to the hypothesized tradeoff: enacting a strong commitment strategy is associated with greater tolerance for experimentation but less growth on the extensive margin. One concern with these results is that they cannot be interpreted causally. Although individual fixed effects help control for time invariant heterogeneity, they are unable to eliminate other sources of endogeneity. Furthermore, I do not directly observe the hypothesized mechanisms. Further work is required to address these empirical issues.

DISCUSSION

Entrepreneurial experimentation is traditionally thought to improve outcomes because it avoids costly commitment and creates real options. Recent work, however, demonstrates that experimentation often entails partial commitments and as a result, experimentation may result in worse outcomes (Chavda 2018; Gans, Stern, and Wu 2019). The present work extends this core insight by highlighting the role of the audience in entrepreneurial experimentation. When audience intolerance to participation threatens the effectiveness of entrepreneurial experimentation, entrepreneurs face a tradeoff between two types of commitment: sunk-cost investments in developing a better prototype versus committing to a relationship predicated on joint value creation. I find that audiences are less likely to penalize entrepreneurs who display credible commitment to them, but that these entrepreneurs are subsequently constrained from pivoting to more attractive markets than entrepreneurs who are not committed to their audience.

More generally, the present work advances our understanding of entrepreneurial strategy and experimentation in two respects. First, the present work highlights why entrepreneurial experimentation must often be regarded as a relational process. Second, the present work extends and challenges an increasingly popular view of entrepreneurs that sees them as quasi-scientists.

Entrepreneurial experimentation as a relational process

Both academic and practitioner-oriented work on entrepreneurial experimentation (e.g., Murray and Tripsas 2004; Contigiani and Levinthal 2019; Felin et al. 2020; Brown 2008; Ries 2011) often references the importance of interacting with end users in learning from experimentation, but this work rarely addresses the unintended impacts of the entrepreneur's relationship to their audience on the dynamics of experimentation.[27] A core insight of the present work is that entrepreneurial experimentation is often a relational process. The more entrepreneurial experimentation relies on cooperation with an audience participants to create joint value, the more important relationships become to successful outcomes.

More generally, the present work contributes to our understanding of entrepreneurial strategy and experimentation by emphasizing the importance of demand-side (audience) factors. By contrast, prior work on constraints on entrepreneurial experimentation has typically focused on supply-side factors, such as the entrepreneur's resources or strategies (e.g., Kerr, Nanda, and Rhodes-Kropf 2014; McDonald and Eisenhardt 2020).

[27] Contigiani and Levinthal (2019) as well as Felin et al. (2020) draw connections between the literature on entrepreneurial experimentation and foundational academic ideas such as organizational learning. To be sure, various challenges associated with learning, especially in the context of embedded relationships (e.g., Anjos and Reagans 2013; Uzzi and Lancaster 2003), have been explored in the broader literature.

The entrepreneur as quasi-scientist

An increasingly popular perspective on entrepreneurial activity casts entrepreneurs as quasi-scientists, who develop theories and test them through the process of experimentation (e.g., Felin and Zenger 2009; 2017; Camuffo et al. 2020). In some cases, the metaphor is used in a positive sense, to suggest that entrepreneurs behave as if they are scientists, whereas in other cases it is used in a normative sense, to suggest that the scientific process represents an ideal that entrepreneurs should aspire to. However, the present work suggests that entrepreneurial experimentation is unlike scientific experimentation in important respects, and that following the metaphor too closely might suggest practices that are in fact detrimental to entrepreneurial outcomes.

A consummate physicist who studies atoms approaches them with a passionate disinterestedness. The atoms certainly have no regard for the physicist—they do not get concerned if the physicist looks at other atoms, nor do they care which physicist studies them. By contrast, entrepreneurs study audiences. These audiences are simultaneously subject and collaborator in the experimental process. At the end of the day, when entrepreneurs and audiences cooperate for joint value creation, they have a personal interest in one another. Yet this view of entrepreneurship as a self-serving and relational process violates nearly all the Mertonian scientific norms in one way or another.

One reason why entrepreneurial experimentation cannot and should not aspire to the ideals of positivistic science is rooted in the nature of the relationship between the experimenter and their audience. In scientific experimentation, experimenters who engage too closely with their participants are typically thought to present threats to internal validity because of experimenter effects (Rosenthal 2005). The solution is a cool, arm's-length relation between participant and experimenter, the gold standard of which is the double-blind randomized control trial. However, as the present work demonstrates, entrepreneurial experimentation can in fact benefit from a close relationship with the audience and may not even be possible without such a relationship.

A second reason why entrepreneurial experimentation differs from positivistic scientific ideals concerns the motivation of the entrepreneur, which is primarily value creation and capture. Approaching entrepreneurial experimentation with a "passionate disinterestedness" would be antithetical to the goals of the entrepreneur. The notion that entrepreneurial experimentation can be universal or objective is particularly problematic in markets where valuation is heavily contingent on producer identity, such as in cultural markets. As a wealth of research by economic sociologists and organization theorists has demonstrated, audiences are sensitive to the identities of producers. A high-status producer may be allowed to deviate from certain norms, whereas such leniency may not be afforded to a middle-status producer (Phillips and Zuckerman 2001). Similarly, the ability of an entrepreneur to experiment and the results they achieve are not universal, but depend on the identity of the entrepreneur.

Limitations

The analyses in the present work face some limitations. In particular, my empirical approach does not allow me to make causal claims. As with all strategy research, any observed relationship between strategy and performance is plagued by endogeneity. A common confounder is unobservable heterogeneity, such as skill or personality, which can be expected to operate in this setting. By using individual fixed effects, I am able to control for time-invariant individual heterogeneity, but cannot rule out time-varying heterogeneity. Even controlling for time-varying heterogeneity, the choice of which new games to experiment with and the timing of adoption would be endogenous.

CONCLUSIONS

This paper contributes to our understanding of the relationship between experimentation and commitment. Whereas recent work documents the ways in which experiments entail partial commitments, this paper documents a setting where audiences do not tolerate experimentation until the entrepreneur demonstrates commitment to them. This presents a tradeoff for entrepreneurs seeking to use experimentation to test the validity of their ideas. On the one hand, commitment enables experimentation because audiences are willing to endure cruder prototypes and provide richer feedback. On the other hand, the act of commitment curtails the switching option and renders moot potential benefits from identifying more attractive markets typically associated with experimentation. Thus, rather than commitment being an (often unintended) byproduct of experimentation, this paper demonstrates that commitment can be a prerequisite for experimentation. While this may not be the case in all settings, at least one setting in which these concerns are likely to be salient include market contexts where producer identity is salient for consumers.

REFERENCES

6000j. July 11, 2021. *Reddit.* https://www.reddit.com/r/speedrun/comments/ohzzvp/which_video_game_and_category_take_the_most_skill/h4se5lt/. Accessed March 21, 2022.

Adams, E. and Dormans, J., 2012. *Game mechanics: advanced game design.* New Riders.

Agrawal, Ajay, Joshua S. Gans, and Scott Stern. 2021. "Enabling Entrepreneurial Choice." *Management Science*.

American Heritage Dictionary of the English Language, Fifth Edition. 2011. Houghton Mifflin Harcourt Publishing Company. Retrieved March 19 2022 from https://www.thefreedictionary.com/mind+game.

Anjos, Fernando, and Ray Reagans. 2013. "Commitment, Learning, and Alliance Performance: A Formal Analysis Using an Agent-Based Network Formation Model." *The Journal of Mathematical Sociology* 37(1):1–23.

Baym, Nancy K. 2018. Playing to the Crowd. New York: New York University Press.

Blight, Michael G. 2018. "Relationships to Video Game Streamers: Examining Gratifications, Parasocial Relationships, Fandom, and Community Affiliation Online."

Brealey, Richard A. and Stewart C. Myers. 2000. Principles of Corporate Finance (6th edn.). New York: McGraw-Hill.

Brown, Tim. 2008. "Design Thinking." Harvard Business Review, June.

Camuffo, Arnaldo, Alessandro Cordova, Alfonso Gambardella, and Chiara Spina. 2020. "A Scientific Approach to Entrepreneurial Decision Making: Evidence from a Randomized Control Trial." *Management Science* 66(2):564–86.

Carroll, Glenn R., and Anand Swaminathan. 2000. "Why the Microbrewery Movement? Organizational Dynamics of Resource Partitioning in the US Brewing Industry." *American Journal of Sociology* 106(3):715–62.

Chavda, Ankur. 2018. "The Downside of Experimentation: Evidence from Television Shows." Job Market Paper.

Clark, Taylor. 2017. "How to Get Rich Playing Video Games Online." November 20, 1–14.

Contigiani, Andrea, and Daniel A. Levinthal. 2019. "Situating the Construct of Lean Start-up: Adjacent Conversations and Possible Future Directions." *Industrial and Corporate Change* 28(3):551–64.

Davis, Gerald F., Kristina A. Diekmann, and Catherine H. Tinsley. 1994. "The Decline and Fall of the Conglomerate Firm in the 1980s: The Deinstitutionalization of an Organizational Form." *American Sociological Review* 59(4):547-70.

Dixit, Avinash. 1980. "The Role of Investment in Entry-Deterrence." *The Economic Journal* 90(357):95.

Durand, Rodolphe and Lionel Paolella. 2013. "Category Stretching: Reorienting Research on Categories in Strategy, Entrepreneurship, and Organization Theory." *Journal of Management Studies* 50(6):1100–1123.

Eisenmann, Thomas, Eric Ries, and Sarah Dillard. 7AD. *Hypothesis-Driven Entrepreneurship: The Lean Startup*.

Felin, Teppo, Alfonso Gambardella, Scott Stern, and Todd Zenger. 2020. "Lean Startup and the Business Model: Experimentation Revisited." Long Range Planning 53(4):101889. doi: 10.1016/j.lrp.2019.06.002.

Felin, Teppo, and Todd R. Zenger. 2009. “Entrepreneurs as Theorists: On the Origins of Collective Beliefs and Novel Strategies.” Strategic Entrepreneurship Journal 3(2):127–46. doi: 10.1002/sej.67.

Felin, Teppo, and Todd R. Zenger. 2017. “The Theory-Based View: Economic Actors as Theorists.” Strategy Science 2(4):258–71. doi: 10.1287/stsc.2017.0048.

Firth, John Rupert. n.d. A Synopsis of Linguistic Theory, 1930-1955.

Frake, Justin. 2017. “Selling Out: The Inauthenticity Discount in the Craft Beer Industry.” *Management Science* 63(11):3930–43.

Freeman, J. and Hannan, M.T., 1983. Niche width and the dynamics of organizational populations. *American Journal of Sociology, 88*(6), pp.1116-1145.

Gans, Joshua S., Scott Stern, and Jane Wu. 2019. “Foundations of Entrepreneurial Strategy.” Strategic Management Journal 40(5):736–56. doi: 10.1002/smj.3010.

Gans, Joshua S., Michael Kearney, Erin L. Scott, and Scott Stern. 2021. “Choosing Technology: An Entrepreneurial Strategy Approach.” *Strategy Science* 6(1):39–53.

Garg, N., Schiebinger, L., Jurafsky, D. and Zou, J., 2018. Word embeddings quantify 100 years of gender and ethnic stereotypes. *Proceedings of the National Academy of Sciences, 115*(16), pp.E3635-E3644.

Ghemawat, Pankaj. 1991. *Commitment: The Dynamic of Strategy*. New York: The Free Press.

Gittleson, Kim. 2014. “Amazon Buys Video-Game Streaming Site Twitch.” *BBC News*. Retrieved September 1, 2021 (https://www.bbc.com/news/technology-28930781).

Granqvist, Nina, Stine Grodal, and Jennifer L. Woolley. 2013. “Hedging Your Bets: Explaining Executives’ Market Labeling Strategies in Nanotechnology.” *Organization Science* 24(2):395–413.

Grayson, Nathan. Ninja Takes Two-Day Break, Loses 40,000 Subscribers. Retrieved April 5, 2022 (https://kotaku.com/ninja-takes-two-day-break-loses-40-000-subscribers-1826813300).

Hahl, Oliver, and Jaekyung Ha. 2020. “Committed Diversification: Why Authenticity Insulates Against Penalties for Diversification.” *Organization Science* 31(1):1–22.

Hannan, Michael T. 2010. “Partiality of Memberships in Categories and Audiences.” *Annual Review of Sociology* 36(1):159–81.

Hannan, M.T. and Freeman, J., 1977. The population ecology of organizations. *American Journal of Sociology, 82*(5), pp.929-964.

Hannan, Michael T., László Pólos, and Glenn R. Carroll. n.d. *Logics of Organization Theory: Audiences, Codes, and Ecologies*. Princeton, NJ: Princeton University Press.

Hayek, F. A. 2002. “Competition as a Discovery Procedure.” *Quarterly Journal of Austrian Economics* 5(3).

Heitner, Darren. 2018. Esports Legend Ninja Confirms He Is Earning Over $500,000 Per Month Retrieved April 5, 2022 (https://www.forbes.com/sites/darrenheitner/2018/03/20/esports-legend-ninja-confirms-he-is-earning-over-500000-per-month/).

Hernandez, Patricia. 2018. “The Twitch Streamers Who Spend Years Broadcasting to No One.” The Verge. Retrieved August 31, 2021 (https://www.theverge.com/2018/7/16/17569520/twitch-streamers-zero-viewers-motivation-community).

Higginson, Matt, Marie-Claude Nadeau, and Kausik Rajgopal. 2019. “Blockchain’s Occam problem.” McKinsey & Company. Retrieved September 6, 2021 (https://www.mckinsey.com/industries/financial-services/our-insights/blockchains-occam-problem).

Hsu, Greta, Michael T. Hannan, and Özgecan Koçak. 2009. "Multiple Category Memberships in Markets: An Integrative Theory and Two Empirical Tests." *American Sociological Review* 74(1):150–69.

Hsu, Greta and Michael T. Hannan. 2005. "Identities, Genres, and Organizational Forms." *Organization Science* 16(5):474–90.

Hsu, Greta. and Stine Grodal. 2015. "Category Taken-for-Grantedness as a Strategic Opportunity: The Case of Light Cigarettes, 1964 to 1993." *American Sociological Review*, 80(1):28–62.

Kerr, William R., Ramana Nanda, and Matthew Rhodes-Kropf. 2014. "Entrepreneurship as Experimentation." *Journal of Economic Perspectives* 28(3):25–48.

Kirzner, Israel M. 1997. "Entrepreneurial Discovery and the Competitive Market Process: An Austrian Approach." *Journal of Economic Literature* 35(1):60–85.

Kovács, Balázs and Michael T. Hannan. 2015. "Conceptual Spaces and the Consequences of Category Spanning." *Sociological Science* 2:252–86.

Leung, Ming D. 2014. "Dilettante or Renaissance Person? How the Order of Job Experiences Affects Hiring in an External Labor Market ." *American Sociological Review* 79(1):136–58.

Leung, Ming D. and Amanda J. Sharkey. 2014. "Out of Sight, Out of Mind? Evidence of Perceptual Factors in the Multiple-Category Discount." *Organization Science* 25(1):171–84.

Manso, Gustavo. 2011. "Motivating Innovation." *The Journal of Finance* 66(5):1823–60.

Manso, Gustavo. 2016. "Experimentation and the Returns to Entrepreneurship." *The Review of Financial Studies* 29(9):2319–40.

Margolis, Eric. 1994. "A Reassessment of the Shift from the Classical Theory of Concepts to Prototype Theory." *Cognition* 51(1):73–89.

Murray, Fiona, and Mary Tripsas. 2004. "The Exploratory Processes of Entrepreneurial Firms: The Role of Purposeful Experimentation." *Advances in Strategic Management* 45–75.

Matsen, F.A., Billey, S.C., Kas, A. and Konvalinka, M., 2016. Tanglegrams: a reduction tool for mathematical phylogenetics. *IEEE/ACM transactions on computational biology and bioinformatics, 15*(1), pp.343-349.

May, E. 2021a. *Streamlabs & Stream Hatchet Q1 2021 Live Streaming Industry Report*. Retrieved September 1, 2021 (https://blog.streamlabs.com/streamlabs-stream-hatchet-q1-2021-live-streaming-industry-report-eaba2143f492).

May, E. 2021b. *Streamlabs and Stream Hatchet Q4 Live Streaming Industry Report.* Retrieved April 9, 2022 (https://blog.streamlabs.com/streamlabs-and-stream-hatchet-q4-live-streaming-industry-report-a898c98e73f1).

May, E. 2022. *Streamlabs and Stream Hatchet Q1 2022 Live Streaming Industry Report*. Retrieved May 5, 2022 (https://streamlabs.com/content-hub/post/streamlabs-and-stream-hatchet-q1-2022-live-streaming-industry-report).

McDonald, Rory M., and Kathleen M. Eisenhardt. 2020. "Parallel Play: Startups, Nascent Markets, and Effective Business-Model Design." Administrative Science Quarterly 65(2):483–523. doi: 10.1177/0001839219852349.

Medin, Douglas L. and William D. Wattenmaker. 1987. "Family Resemblance, Conceptual Cohesiveness, and Category Construction." *Cognitive Psychology* 19(2):242–79.

*Melee has the hardest learning curve (*author unknown). May 22, 2017. *Reddit.* https://www.reddit.com/r/esports/comments/6cnwj9/what_are_the_top_5_games_that_have_the_highest/dhwfduv/. Accessed March 21, 2022.

Merluzzi, Jennifer and Damon J. Phillips. 2016. "The Specialist Discount: Negative Returns for MBAs with Focused Profiles in Investment Banking." *Administrative Science Quarterly* 61(1):87–124.

Murphy, Gregory L. and Douglas L. Medin. 1985. "The Role of Theories in Conceptual Coherence." *Psychological Review* 92(3):289–316.

Nanda, Ramana, and Matthew Rhodes-Kropf. 2016. "Financing Entrepreneurial Experimentation." *Innovation Policy and the Economy* 16:1–23.

National Institute of Standards and Technology. 1996. *DATAPLOT Reference Manual,* pp. 2-66. https://www.itl.nist.gov/div898/software/dataplot/refman2/ch2/weightsd.pdf. Accessed March 21, 2022.

Negro, Giacomo, Özgecan Koçak, and Greta Hsu. 2010. *Research on Categories in the Sociology of Organizations*. Elsevier.

Negro, Giacomo and Ming D. Leung. 2013. "'Actual' and Perceptual Effects of Category Spanning." *Organization Science* 24(3):684–96.

Osherson, Daniel N. and Edward E. Smith. 1981. "On the Adequacy of Prototype Theory as a Theory of Concepts." *Cognition* 9(1):35–58.

Paolella, Lionel and Rodolphe Durand. 2016. "Category Spanning, Evaluation, and Performance: Revised Theory and Test on the Corporate Law Market." *Academy of Management Journal* 59(1):330–51.

Phillips, Damon J., Catherine J. Turco, and Ezra W. Zuckerman. 2013. "Betrayal as Market Barrier: Identity-Based Limits to Diversification Among High-Status Corporate Law Firms." *American Journal of Sociology* 118(4):1023–54.

Phillips, Damon J. and Ezra W. Zuckerman. 2001. "Middle-Status Conformity: Theoretical Restatement and Empirical Demonstration in Two Markets." *American Journal of Sociology* 107(2):379–429.

Reagans, Ray E., and Bill McEvily. 2003. "Network Structure and Knowledge Transfer: The Effects of Cohesion and Range." *Administrative Science Quarterly* 48(2):240.

Reagans, R.E. and Zuckerman, E.W., 2008. All in the family: reply to Burt, Podolny, and van de Rijt, Ban, and Sarkar. *Industrial and Corporate Change, 17*(5), pp.979-999.

Retro Ali. 2017, March 16. Single Game VS Variety Streaming | Twitch Tips #23 [Video]. YouTube. Retrieved August 31, 2021 (https://www.youtube.com/watch?v=lDD3UDMv0ks).

Ries, Eric. 2011. *The Lean Startup*. New York, NY: Crown Business.

Rosch, Eleanor. 1978. "Principals of Categorization." Pp. 1–23 in *Cognition and Categorization*.

Rosch, Eleanor. 1975. "Cognitive Representations of Semantic Categories." Journal of Experimental Psychology: General 104(3):192–233.

Rosch, Eleanor and Carolyn B. Mervis. 1975. "Family Resemblances: Studies in the Internal Structure of Categories." *Cognitive Psychology* 7(4):573–605.

Rosenthal, Robert. 2005. "Experimenter Effects" edited by K. Kempf-Leonard. Encyclopedia of Social Measurement 871–75.

sadpandag. 2015, December 15. Variety Streaming is Tough: Exploring the State of Variety Broadcasting on Twitch [Online forum post]. Reddit. Retrieved August 31, 2021 (https://www.reddit.com/r/Twitch/comments/3wywyk/variety_streaming_is_tough_exploring_the_state_of/).

Scornavacca, C., Zickmann, F. and Huson, D.H., 2011. Tanglegrams for rooted phylogenetic trees and networks. *Bioinformatics, 27*(13), pp.i248-i256.

Sgourev, Stoyan V. and Niek Althuizen. 2014. "“Notable” or “Not Able”." *American Sociological Review* 79(2):282–302.

Smith, Edward E. and Daniel N. Osherson. 1984. “Conceptual Combination with Prototype Concepts.” *Cognitive Science*.

Soulcloset. April 3, 2019. *Reddit.* https://www.reddit.com/r/speedrun/comments/b94g4s/what_do_speedrunners_think_of_rng/ek24i1m/. Accessed March 21, 2022.

Statista. 2015. *Video game retail sales value in the United States from 2000 to 2015.* Retrieved April 3, 2022 (https://www.statista.com /statistics/530168/total-revenue-of-us-video-games-market/).

Statista. 2022. *Registered users of Fortnite worldwide from August 2017 to May 2020*. Retrieved April 5, 2022 (https://www.statista.com/statistics/746230/fortnite-players/).

SullyGnome. n.d. “Ninja.” Retrieved May 5, 2022 (https://sullygnome.com/channel/ninja/2018/compare).

Taylor, T. L. 2018. *Watch Me Play.* Princeton, NJ: Princeton University Press.

Techcrunch. 2007. Justin.TV Teams Up With On2 And Opens Network. https://techcrunch.com/2007/10/02/justintv-teams-up-with-on2-and-opens-network-finally/. Accessed April 4, 2022.

Trigeorgis, Lenos, and Jeffrey J. Reuer. 2017. “Real Options Theory in Strategic Management.” *Strategic Management Journal* 38(1):42–63.

Twitch. n.d. “Achievements.” Retrieved May 4, 2022 (https://help.twitch.tv/s/article/achievements).

Twitch. n.d. “Joining the Affiliate Program.” Retrieved May 5, 2022 (https://help.twitch.tv/s/article/joining-the-affiliate-program).

Twitch. n.d. “Twitch Partner Program.” Retrieved May 5, 2022 (https://www.twitch.tv/p/en/partners/).

Twitch Blog (blog). 2021. “Let’s Talk About Hot Tub Streams.” Retrieved May 5, 2022 (https://blog.twitch.tv/en/2021/05/21/lets-talk-about-hot-tub-streams/).

Twitch Creator Camp. 2018. “Level Up - Finding Your Niche on Twitch w/ SimCopter1.” Retrieved on October 31, 2021 (https://www.twitch.tv/videos/293880183).

Uzzi, Brian, and Ryon Lancaster. 2003. “Relational Embeddedness and Learning: The Case of Bank Loan Managers and Their Clients.” *Management Science* 49(4):383–99.

Vergne, J. P., and Tyler Wry. 2013. “Categorizing Categorization Research: Review, Integration, and Future Directions.” Journal of Management Studies 51(1):56–94.

von Hippel, Eric. 2007. “Horizontal Innovation Networks—by and for Users.” *Industrial and Corporate Change* 16(2):293–315.

Webb, Amy. 2016. “The 'Uber for X' Fad Will Pass Because Only Uber Is Uber.” *Wired*. Retrieved September 6, 2021 (www.wired.com/2016/12/uber-x-fad-will-pass-uber-uber/).

Zuckerman, Ezra W. 1999. “The Categorical Imperative: Securities Analysts and the Illegitimacy Discount.” *American Journal of Sociology* 104(5):1398–1438.

Zuckerman, Ezra W. 2000. “Focusing the Corporate Product: Securities Analysts and De-Diversification.” *Administrative Science Quarterly* 45(3):591–619.

Zuckerman, Ezra W. 2004. “Structural Incoherence and Stock Market Activity.” *American Sociological Review* 69(3):405–32.
Zuckerman, Ezra W., Tai Young Kim, Kalinda Ukanwa, and James von Rittmann. 2003. “Robust Identities or Nonentities? Typecasting in the Feature-Film Labor Market.” *American Journal of Sociology* 108(5):1018–73.
Zuckerman, Ezra W. 2017. “The Categorical Imperative Revisited: Implications of Categorization as a Theoretical Tool.” Pp. 31–68 in *Research in the Sociology of Organizations*, vol. 51. Emerald Group Publishing Limited.