Master's Thesis

Relationships Between Complementing Platforms

A quantitative case study on factors influencing cross-platform behavior

Names (Student number): Kasper Steffensen (101651) & Lennart Zellmer (124438)

Educational Program: Business Administration and E-business (EBUSS)

Supervisor(s): Philipp Hukal & Irfan Kanat

Number of characters: 262064

Number of normal pages: 115

Number of physical pages: 114

Date: 15/05/2020

Exam: Master's Thesis (CBUSO2000E) - Kontraktnr: 16756

Abstract

The presented paper examines the interconnection of digital platforms and answers the following research question: ***Which factors influence cross-platform interaction from users on complementing platforms?*** In our conducted literature review, we were able to abstract the concept of complementing platforms from the literature on complementing products and platform complementors. This concept of complementing platforms is directly connected to the platform's horizontal or vertical position within platform networks.

We attempted to answer this question through a quantitative big data approach, by collecting data from two complementary platforms, Twitch and Steam. At the data collection stage, we observed the user behaviour on the two platforms over one month through appropriate resource boundaries. Thereby we were able to obtain a comprehensive data set on Twitch viewer- and Steam player activity and interaction. We then tested the hypotheses developed in the research design section and used multiple linear regression to test the hypotheses on the users of each platform. The results were that all variables were statistically significant except for hardware requirements that affected the players of Steam. The statistically significant results showed that a **progressing product life cycle** influences cross-platform interaction in a positively correlated way on Steam, and a negatively correlated way on Twitch, which led us to find that there are two classes of games on the platforms. The first class is what we define as the traditional life cycle game and the second as a multi-player service game with a surrounding game framework most closely related to a traditional sport. **Social interaction** also influenced cross-platform interaction, in that it increased interaction on both platforms, confirming our initial expectations and prior research on the topic. Finally, we found that **access barriers** influence cross-platform interaction from users on complementary platforms and that **access barriers** act as an *access barrier* on one platform, but as an *access facilitator* on its complementing platform.

Keywords: Platforms, complementors, platform ecosystems, video-game industry, platform coupling

Table of Contents

List of tables

List of figures

Introduction

It is fascinating to observe, and almost appears unreal, which breakthroughs and innovations have been necessary to create the business models of digital platforms. In the last few years and still today, we experience a development that is unique and unprecedented in the history of humanity. Our opportunities for research and inventiveness have taken on exponential proportions based on both technological and sociological factors.

Starting with the development of consumer-ready computers in the 1990s, through the rise of the dotcom bubble and associated market development, it has already been a difficult transition for the platform industry and the digitalization of two-sided markets. Due to exponentially smaller device sizes during the 2000' reaching its breakthrough with the introduction of the iPhone and revolutionizing the mobile market, computers and thus digital solutions became suitable for large-scale mass-markets. Through the combined influence of further exponential effects such as the increase in computing capacity according to Moore's Law or the cost reduction in storage technology, a far-reaching saturation with technically highly advanced, constantly available digital mobile devices was achieved in less than a decade. All these individual technological and social elements generate multiplier effects, resulting in the emergence of very successful, scalable as well as profitable business models. Eight of the ten most successful international companies rely on multiple platforms as a critical aspect of value creation.

This increasing economic and societal relevance of platforms, which today affect almost every aspect of our daily lives, also fuelled the scientific interest in this area since the early 2000s. Building on the intensive research on two-sided markets and software architecture, a comprehensive academic discussion of the technological-, market- and user-based aspects of platforms developed. The determination to explain and understand platform dynamics as accurately as possible is evident in many studies that illuminate the dynamics and effects within platforms itself. The combined perspectives of economic theory and software architectural engineering were able to theoretically explore and conceptually understand network effects, complementors, resource boundaries, modularity, and other platform defining concepts.

While research on dynamics and market behaviour within a platform seem to have significantly benefited from the high level of interest generated, the area of research regarding relationships among platforms in ecosystems is less developed and researched. Fundamental ideas about platform ecosystems and macro-level dynamics are not yet accompanied by the consideration of actual relationships between individual platforms. Therefore our thesis aims to understand the interaction of platforms holistically with focus specifically on the particular interlinking of such platforms.

Looking at these large multi-platform corporations, it becomes evident that in most cases complementary platforms extend the offer of a main application. At Facebook, for example, differentiated offers such as Instagram or Whatsapp set complementary counterparts to the central platform of the same name. Also, at Alphabet Inc., as holding company combines complementary platforms such as Google Search, Android, Google play store and YouTube. Therefore, we will focus our thesis, especially on such complementing platforms as they unfold both scientific relevance through gap in research as well as practical relevance based on concrete cases of application.

A unique market characterized by transparency and openness, making it particularly interesting for observing such relationships, can be found in the video game industry. Over the last three decades, a diverse ecosystem has developed that operates as a predominantly complementary network with participants such as Xbox, Youtube Gaming, Epic Games Store, Mixer, Discord and numerous other parties. Besides the academic and practical relevance of the topic, we are also inspired by a profound personal interest within the gaming field. We have both grown up with video games as a part of our life, and we have seen the rise of the industry first hand. We hope to shed some light on the dynamic state of the industry through this paper and uncover the relationship between two complementing platforms. From our own experience, we can state that the user perspective strongly influences the relationship between platforms in the video game industry. The inherently very digital target group of relatively young consumers leads to an interesting user-centric approach. Complementary characteristics of platforms may also only be experienced by users and must, therefore - also within this thesis - be considered from this very perspective.

As one of the first attempts to examine the actual connections of platforms, we intend to identify potential, influential factors that influence such a complementary relationship. We will, furthermore, evaluate their impact applying a quantitative case study of the complementary platforms Twitch and Steam. These considerations lead to the following research question:

Which factors influence cross-platform interaction from users on complementing platforms?

Literature Review

The platform research is a wide-ranging and diverse domain that combines and interlinks numerous technical and economic factors. In the following, we intend to relate this diverse field to the specific research question and highlight relevant aspects of existing research projects.

In order to do so, it is essential to first clarify the meaning of the term "digital platforms" and how it emerged to then discuss their distinctive features and characteristics that set them apart from traditional business models.

State of platform research

The research on the phenomenon of platforms can be categorized mainly into two dimensions, a technically driven view of information systems research and a market-based view that focuses more on the organization, innovation and market power of multi-sided markets. The term of *digital* platforms, which has been discussed with increasing intensity since the early 2000s, can in a way be seen as the result of the convergence of these two research branches.

A starting point for research on this topic can be found in the academic discourse concerning two-sided markets that originated since the early 1980s. Even though the specific idea of platforms, let alone *digital* platforms, had not yet emerged, Michael L. Katz and Carl Shapiro already describe the effects surrounding two-sided markets (Katz & Shapiro, 1985). By connecting two or multiple user-groups platforms can leverage those network effects, for rapid growth and utility potential. In this constellation, an increasing number of users also provides for increased usefulness of the overall system for its respective consumers. Under certain conditions, it is possible to create positive feedback loops which are especially relevant to unfold winner-take-all dynamics (Arthur, 1989; Shapiro et al., 1998, p. 299). An example of such positive network effects can be seen, especially on social media platforms today. In this case, direct network effects can be observed as the same user group increases their utility because other users sign up to use a platform's services.

This first non-technical examination of network markets received increased interest in the 1990s, by considering specifically the distribution of market power in such circumstances. The research of Nobel Prize winner Jean Tirole and Jean-Charles Rochet provided the foundation for understanding this market power and price-setting strategies in two-sided markets (Rochet & Tirole, 2003). By analyzing payment service institutions such as credit card providers, the authors were the first to significantly shape the terminology of multi-sided and two-sided markets as platforms. In addition, by conducting mini case studies, the authors were able to identify numerous industries as two- and multi-sided markets. Among the studied credit and debit card providers, and operating systems, they already considered streaming-media technology as well as video games to be suitable for the discussed dynamics. In the following years, this field of research was extended by various additional industries, such as the health care sector, or an even more detailed analysis of payment service providers. (Eisenmann et al., 2006; David Sparks Evans & Schmalensee, 2005)

All these efforts to better understand the phenomenon of two-sided markets were inspired in particular by economic theory but left aside the technical possibilities and developments that have taken place over the last thirty years. Therefore, in the early 2000s, a technology-driven scientific community developed which perceived platforms mainly as "modular architectures" (Baldwin et al., 2009).

This technical perspective on platforms is on the one hand based on product development strategies, in which modularity allows for the reduction of development costs and faster product innovation cycles (See Muffatto & Roveda, 2000). On the other hand, its roots can be found in software development processes and the architecture of large and complex computer programs. Here, standardized interfaces and a well-defined hierarchy provide for the possibility to redefine the functionality of systems retrospectively (Yoo et al., 2010).

These subsequent changes to the functionality can be illustrated by the example of the smartphone operating systems. Here, after the release and sale of a device, the purpose and usage can be modified entirely by readjusting the application layer. On the particular basis of Android, for example, an app developer can map a wide variety of applications on the identical underlying software platform (Svahn & Henfridsson, 2012, p. 3352). If one examines the example even more closely, one can easily determine how the individual technical characteristics of platforms interact. De Reuver et al. summarize this process in which "app developers combine existing layered-modular resources from the operating systems, the various hardware elements, the software development kits and a variety of public application programming interfaces (APIs) into novel apps not considered when the smartphones and associated software were conceived" (2018, p. 3)

Gawer concludes that the technical perspective perceives platforms as "purposefully designed technological architectures (including interfaces) that facilitate innovation." (2014, p. 1243)

Considering the research question on interaction effects of two complementary platforms, a less technical approach focusing on market effects between two participants appears to be reasonable at first sight. However, if we consider the rapid rate of technological innovation that restructures such market conditions, it also seems necessary to discuss the technical architectural aspects of platforms. It, therefore, seems appropriate to apply an integrative approach addressing the research question that combines both the market economy and technical engineering perspectives.

In her attempt to bridge the two opposing views of economics and engineering, Gawer (2014, p. 1245) develops a unified integrative framework that manages to combine the seemingly opposing platform descriptions. She strives to capture the diversity of the platform in the actual environment and to design the framework independently of the organizational context of a platform. Furthermore, the framework should allow multimodal interaction between platform agents, either within or across platforms. Especially the latter is in its definition essential for answering our research question.

Gawer summarizes her unified conceptualization as follows: "Technological platforms can be usefully seen as *evolving organizations* or *meta-organizations* that: (1) federate and coordinate constitutive agents who can innovate and compete; (2) create value by generating and harnessing economies of scope in supply or/and in demand; and (3) entail a technological architecture that is modular and composed of a core and a periphery."

We adopt this definition of platforms for the present work because the case of complementary platforms may also include the combination of a diverse field of organizations.

In this context, it is quite possible that a high-tech industry platform might be complementary to a consumer-oriented platform. An example of this would be the hosting service Digital Ocean and the Amazon Web Services (AWS) which in a way, offer complementary services, but provide them in other markets through very different distribution channels. While AWS provides highly specific cloud applications individually, Digital Ocean bundles them into directly operational environments. The two organizations being complementary, should both be considered as platforms although they offer very different products in different markets with different distribution models on a different technology stack.

The classification of platforms proposed by Gawer (2014) allows for this flexibility and is therefore suitable as a definition of platforms used in the following research.

Cross-Platform Effects and Effects between two platforms

Further concepts that require delimitation and clarification when considering the research question are the terms of cross-platform effects and platform networks. In the following section, we describe why cross-platform effects in the strict definition and the particular case are not suitable for the analysis while the concept of platform networks, however, can contribute supportive fundamentals to our study.

While one might assume by name alone that the so-called cross-platform effects exist between two or more platforms and therefore across the boundaries of a platform, current literature describes them as effects that are contained on a solitary multi-sided platform only (Caillaud & Jullien, 2001; D. S. Evans & Schmalensee, 2005; Hoelck et al., 2016; Rochet & Tirole, 2003). These are no different from the network externalities already discussed and can again easily be illustrated using well-known cases. For instance, the more landlords offer their apartments on Airbnb, the more attractive the platform becomes for tourists and vice versa.

This definition of a cross-platform effect differs fundamentally from the interpretation addressed in our research question. Our explicit focus is not on effects that only occur on one platform, but on the relationship between two established platforms.

These connections and linkages between platforms, however, appear to be largely unresearched.

One of the few contributions to this relationship between two or more platforms can be found in the literature on network-centric innovation and innovation ecosystems. It focuses in particular on the way companies, which do not necessarily have to be platforms, can produce innovation through openness and modularity as well as managing the network innovation process. However, the factors of managing such innovation networks can be especially applicable in the context of platforms because they feature a much more open and modular structure due to the inherently digital nature of the business model. However, the management of innovation outside the platform plays a subordinate role concerning our research question regarding the effects between platforms at the consumer level. The network component of this research, on the other hand, poses noticeable parallels to our perception of platforms ecosystems and their complementary peers.

While some studies, like the one of Nambisan and Sawhney (2011), focus specifically on this process of orchestration and exclude the definition of the network and interaction effects, others focus precisely on the conception of this very network.

One of these publications is the one by Adner and Kapoor (2010) as it addresses the creation of value in such innovation ecosystems. For this purpose, the authors develop a generic schema (figure 01) of such ecosystems and describe in particular complementary organizations and suppliers as relevant participants. They thereby focus on the value creation process for the customer and describe a network in which value is created along an upstream procedure. In this process, suppliers deliver modules to a focal firm that assemble those modules to the final product, which is at the end delivered to a customer. While the center of the analysis is formed through one focal firm, the authors acknowledge that customer utility can also be dependent on other complementary products delivered by complementors.

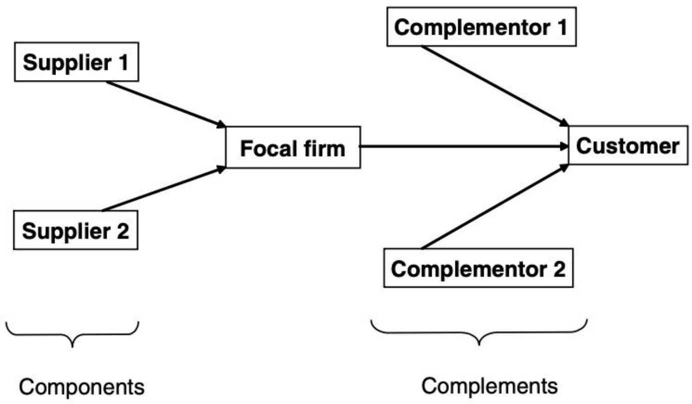

Figure 01. Generic schema of an ecosystem (Adner & Kapoor, 2010)

Adner and Kapoor observe the innovation complexity and competitive advantage of the focal firm in comparison to the upstream suppliers and downstream competitors and their respective innovation complexity. In doing so, they can predict how the focal firm's need for innovation as well as competitive advantage develops depending on the overall ecosystem. Although they develop the idea of a network in which potential interaction between the participants is feasible, they nevertheless separate the individual participants in their analysis according to their function as suppliers and complementor. The analysis is therefore not aimed at the interaction of the focal firm with a complementary partner, but rather describes the actual state of the entire ecosystem.

Furthermore, the directional dimension of downstream and upstream adds too much complexity to the construct to be considered in our analysis. Also, the weak coupling between the two complementors does not meet our perception of the relationship between them.

Even if the basic fit of this model does not match the given research question, two elements of the analysis will be interesting for further consideration.

Firstly, the life cycle of a product or technology could be relevant for the interaction effects of the two platforms on a complimentary level. Adner and Kapoor describe the role of the life cycle as follows: "Early in a technology's life cycle, technological uncertainty is at its peak. As development takes place, knowledge is accumulated, and progress becomes more predictable" (2010, p. 314). Even if, in contrast to the authors, we do not view the ecosystem from the position of a focal firm, but rather from the position of user interaction, the life cycle may also be relevant from this perspective. It is, in fact, quite conceivable that different platforms react differently to technological and product life-cycles. However, this aspect has to be discussed at a later stage because this section will concentrate on the concept of cross-platform effects and the theoretical background of platform interactions.

The second aspect that seems to be reasonable to include in our scenario is the idea of increased customer utility through the consumption of two separate goods. This suggests that a primary link between the platforms can be established mainly through the user and their consumption of two distinct services as platforms. This highlights the fact that the user is in a unique position between the two platforms, to which the two counterparts should relate and align. There are many examples of such a constellation. One has to be careful though to differentiate between multi-homing and complementary relationships in this regard.

Netflix and other streaming service providers are just one of the many possibilities that can illustrate this difference. In this case, many users can only achieve optimal utility for themselves if they are using different platforms like Disney+ or Amazon Video simultaneously (See Park et al., 2018). In principle, Netflix and competing vendors offer the same service with partially divergent content or inventory by giving the user access to TV shows and movies for streaming. Also, with other examples such as Xbox and Playstation show a similar pattern where both products enable the user to play video games with comparable content. These cases and the associated user behaviour has already been largely recognized in recent literature as multi-homing and will therefore not be the focal point of this thesis.

Our research on the relationship between platforms distinguishes from multi-homing in that we require one platform to offer genuine added value to the user on the other platform - ideally reciprocal. In the case of Netflix, for instance, we do not consider the relationship to Amazon Video but could reflect on the user's relationship to the Internet Movie Database (IMDb). This service is actually beneficial to the user experience on Netflix as it provides a great amount of meta-information and reviews on movies and series. Given the fundamental differences between the services offered by IMDb and Netflix, multi-homing in the sense of the platform literature is not possible.

While the example clearly demonstrates the difference between multi-homing and the relationship we want to observe between platforms, it is also able to show another relevant process that Hoelck et al. (2016) describe as coupling. In their work "Competitive Dynamics in the ICT Sector: Strategic Decisions in Platform Ecosystems", the researchers describe the idea of a multidimensional network of actual platforms. Thereby they are the first to introduce specifically platforms into the construct of a network.

Coupling in this network is a condition in which the growth of (products on) one platform also increases the utility of the other platform (figure 02). Speaking in terms of the above example, it seems likely that a larger number of videos and TV shows offered on Netflix will also add value for IMDb users who can generate and review more diverse content and even more vice versa.

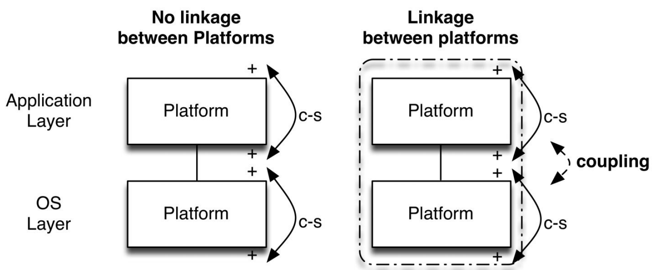

Figure 02. Generic coupling between Platforms (Hoelck & Ballon, 2015)

Through analysis in the ICT sector, the researchers integrate the effect of coupling into a larger construct of phenomena they call platform networks (PNs).

In mature and, in particular, digital platform markets, the platforms can not only exist alongside each other as competitors "but also on top of each other in the value chain creating a complex ecosystem consisting of several layers of platforms." (Hoelck & Ballon, 2015) Therefore platform network can be defined as a multi-layered platform ecosystem which is illustrated in figure 03.

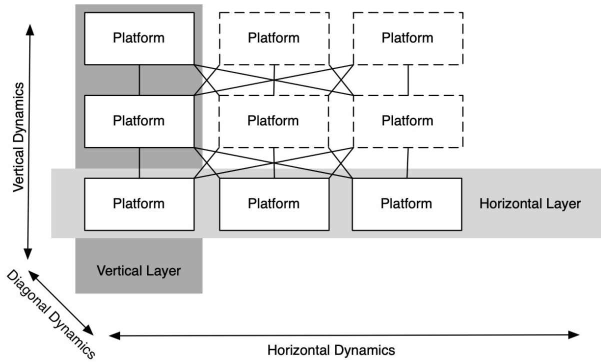

Figure 03. Platform Ecosystem (Hoelck & Ballon, 2015)

While horizontal dynamics describe the competition between companies in the same market offering substitutes, vertical dynamics unfold among complementors within the value chain. Lastly, diagonal dynamics become visible when interaction with outside the primary ecosystem of a platform takes place. (Hoelck et al., 2016, p. 7)

The model allows the representation of complex interactions between platforms which - with an increasing number of actors - are quickly no longer visualizable on two-dimensional paper.

The relationship between the example of Netflix and IMDb, however, can be applied to the above example network quite accurately. While on the horizontal layer video distribution networks such as Netflix, Amazon Video, or Disney+ can be arranged, on the vertical layer platforms such as IMDb, justwatch or VPN-Services may be located. It is also possible to argue that services such as IMDb are better classified as complementors form adjacent markets and are therefore exposed to diagonal dynamics. However, for the analysis presented in this paper, implications are considered insignificant, and therefore diagonal and vertical competition is treated interchangeably.

Complementary Goods and Platforms

Since we now have a good understanding of the context in which the relationship we want to investigate appears, the question arises how this relationship can be described. What does complementarity mean in the context of platforms?

Especially the topic of relationships between platform-based products has been a major area of interest for the scientific community for a long time. Therefore, in the following, we will provide an overview of the discoveries and findings regarding complementary platform-related interactions, with particular focus on the computer and video-game market.

Complementary Goods and Platforms

Had digital platforms and computers existed in 1838, Cournot certainly would not have chosen the manufacturers of zinc and copper to provide one of the first descriptions of complementary relationships. His mathematical economic analysis demonstrates that, in a system where there are only two companies, these producers of raw materials share profits equally, regardless of their marginal costs. (Cournot, 1897, Chapter 9)

While the analyses of complementary products in the last century focused on material goods, since the beginning of the 21st century, the subjects of research are shifting more and more towards digitalized products and services. (See Chen & Nalebuff, 2006; Schilling, 2003)

The examples of applications mentioned by researchers also changed and developed noticeably in recent years from grocery items through tangible computer hardware to service-oriented computer software. (See Yalcin et al., 2013; Yan & Bandyopadhyay, 2011)

Nowadays, we may perceive the products zinc and copper, which Cournot examined in 1838, as hyper-simplistic. Compared to the technological and social efforts required to produce microprocessors and software products, this seems perfectly understandable, yet it also indicates that the scientific ambition to question and explain more complex market situations has intensified over the last 150 years.

However, most of the issues addressed up until today are not approached from a social science perspective regarding user behaviour, but rather examine pricing strategies and cost structures of complementary products. Our work aims to extend this static view on price/cost analysis to include social components of interaction between complementary products.

This research gap can be observed in scientific debate and definitions of complementary goods as well as complementors in the sense of the platform literature. To specify goods or products as complementary arises here from two fundamentally different theoretical approaches: a demand-oriented definition, as well as a consumer utility-based understanding. (Allen, 1934; Hicks & Allen, 1934)

Two of the most prominent representatives of the demand-oriented definition of complementary are the authors Gregory Mankiw and Mark P. Taylor. Their internationally recognized standard reference book "Principles of Macroeconomics" defines complements as "two goods for which an increase in the price of one leads to a decrease in the demand for the other" (2016, p. 70).

Within the standard literature, they are far from being isolated with this interpretation. Also, the textbooks "Microeconomics" (Pindyck & Rubinfeld, 2013) or "Contemporary Economics" (Carbaugh, 2016) follow this concept which is characterized and best visualized by indifference curves and Leontief-functions.

Let us consider the example of the presumably complementary platforms YouTube Gaming (a live stream service for video games) and the Epic Game Store (a platform for the distribution of video games) in the following. In the context of universal standard definition, it seems highly controversial whether these can be considered of complementary nature altogether. However, being comparable to the subject of this study, they would have to be in a complementary relationship in order to answer the research question.

This would require that, as a complementary good, demand for YouTube Gaming would have to decrease if the Epic Game Store increased its prices. However, thinking this thought experiment through raises two questions of understanding and consistency.

1. Can a platform be considered a good in the sense of complementary goods?
2. Does the relationship between price and demand behave as indicated for complements?

First of all, it is debatable whether platforms in the sense of this paper can be considered as (complementary) goods. Examples for those goods provided by Mankiw and others are mostly rather simple commodities such as clothing, food or raw materials. Platforms, however, appear to be much more complex and abstract than simple household products.

The mechanisms of value creation for customers are often not immediately recognizable from the outside as they can be hidden in complex business logic. The services offered by a platform can hardly be divided into similarly simple units as the goods mentioned by Mankiw. Only the interaction of the individual components creates the value of the platform.

This phenomenon can be observed clearly in the example of Airbnb. A simple service comparable in complexity to the commercial goods mentioned by Mankiw would be the rental of vacation apartments. But this is not the service provided by the Airbnb platform. The value for the platform is only generated by reducing transaction costs between tenant and landlord as well as creating a relationship of trust through various measures. This combination and interaction of various service features create a new integrated and combined product that can only be analyzed and evaluated in its final configuration. This combined product is therefore also subject to the basic assumptions of supply and demand as well as the demand-oriented definition of a complementary good.

The second question to be addressed is whether the demand for one platform will decrease if the price of the other platform increases.

To answer this question, we need to determine what the actual price for using the given platforms is. What may sound like a trivial question can be difficult to answer in the field of platforms. The Epic Game Store platform on one side includes the costs of using the service in the game price itself as it charges a 10 per cent commission for a game purchase. These indirect costs for the user are still rather transparent compared to other platforms driven by advertising revenues. Thus, in the case of YouTube, it is not particularly straightforward to identify how an increase in price could be implemented in practice.

But let us assume for the sake of argument that also YouTube would be able to adjust prices, e.g. through the introduction of a monthly base fee.

Does this price increasing measure by Youtube Gaming really lead to reduced consumption on the side of the Epic Game Store?

This is probably not the case, as one would expect the exact opposite effect. Increasing the price on one platform will reduce the use of the platform and allow more time to spend on the other. Also, in the reverse relationship, should the Epic Game Store increase its prices, an increase in consumption of YouTube is more likely to be anticipated. If games should suddenly become more expensive, watching live streams or videos might prove to be a cheaper alternative.

Surprisingly, a completely different conclusion can be drawn if one does not look at the platform as a whole but at the games that appear on it. Should a single game increase in price and its demand decrease as a result, there would probably also be potentially less traffic on related streams or videos. Even the reverse effect can at least be deduced logically. Should Youtube increase its prices for a single game's streams, lower demand for this game on the Store can be reasoned due to reduced advertising effects.

This contradiction is difficult to explain since from a superficial external perspective the two goods would quickly be identified as complementary. Perhaps future research will be able to resolve this contradiction by taking a closer look at cost structures in the attention economy. The complex pricing of modern platforms, as well as the challenging effects in the special case of the video game industry, do not allow for a definitive answer whether platforms - and especially media platforms - can be considered complementary in a traditional definition.

The first tentative assessment that Youtube and Epic Game Store may be considered complementary, however, can be underlined with a brief look at the Oxford Dictionary. Etymologically derived from the Latin word *complēmentum* - which translates to "that which fills up or completes" - the word *complement* today means "the action of fulfilling or completing". (*"Complement,"* 2020)

This rather general definition of a complement can be seen as a drastically simplified version of the utility-based approach. Accordingly, two goods are complementary if they "are utilized in combination with one another. Typically, a complementary good has limited significance when used alone but, when used with its complementary products, its overall utility increases." (Cooper, 2015)

This is the very effect that many players are likely to experience when using Youtube and Epic Game Store. But how exactly can a user derive value from the different platforms and can they determine an optimal equilibrium?

While Epic's defined product of the game itself makes it easy to determine the utility provided by the platform, YouTube Gaming is more complicated. The fundamental question here is, why do people watch other gamers playing in the first place?

In their representative study, Sjöblom and Hamari (2017) identify five motivational dimensions that influence the consumption of gaming live streams (cognitive, affective, social, tension release, and personal integrative). While they identify the social component as the primary driver for watching live streams, so do cognitive processes such as eagerness to learn new strategies and skills. Gros et al. (2017) also identify the search for information on different games as one of the major reasons to watch others play. Sometimes even the dimensions of social and information search overlap when viewers start "communicating with other viewers, as they may have new information as well". The information collected here is in turn used in one' s own gaming experience, e.g. on the Epic Game Store platform. The social interactions discovered on YouTube can also be transferred to the Epic Game Store through multiplayer matches playing actively. Friends made on Youtube Gaming can become friends on the Epic Game Store and vice versa.

From our own personal observation, such behaviour can almost be described as a circular and self-reinforcing process. In this procedure, the two activities alternate to either recover from the other activity or to spend time in a different social context for a while.

Even though we have now discussed an example that is close to the industry which we will examine later, this theoretical reflection is also feasible for many other examples. The same behaviour may be observed between the platforms Airbnb and TripAdvisor, Netflix and IMDb or YouTube and Amazon Marketplace with identical conclusions.

Therefore, in the following thesis, we consider platforms as complementary when used jointly to provide a superordinate good enabling the user to enjoy greater utility.

Platform complementors and complementing platforms

We previously already recognized the need for delimitation of the terms platform complementors and complementary platforms. In the following, we would like to revisit this position in order to discuss it in more detail. Even at the risk of appearing repetitive in this section, the precise delimitation and definition of the concepts is an essential foundation for answering the research question.

In contrast to the term complementary platforms, the concept of platform complementors is not unknown to anyone who studies digital platforms for a longer time. If one observes the described complementary relationship between Epic Game Store and Youtube Gaming as an example, the impression may arise that Youtube Gaming could be a *platform complementor* of the Epic Game Store. Thus, in the following, we aim to clarify that the relationship between the two platforms should not be confused with the concept of complementors on one platform.

Usually, whenever leading publications address platform complementors, researchers define the term as an independent third-party provider offering additional services on a provider's platform under the particular rules and conditions of that platform. More specifically, Gawer & Cusumano (2014b) describe complementaries as any actor creating complementary offers on the platform in the form of products, services, or applications.

Also, Wessel et al. (2017) describe complementarities as actors on an existing platform "who develop and deliver the respective content for the platform (e.g., apps, add-ons, plug-ins, modules, or extensions)." They thereby follow the definition of Ghazawneh and Henfridsson published in (2010). These authors do not refer to the specific concept of complementors but describe the phenomenon as actors who, with the help of boundary resources, develop complementary assets for an existing platform.

All these definitions describe a complementary as a product or content that is substantially dependent on the functionality provided by the platform to generate added value. Conversely, this would also mean that platforms could enforce steering and management instruments against the complementary systems. Both of these aspects do not match the observed case and the relationship between the Epic Store and Youtube. Neither does Youtube's service operate on Epic's platform, nor is there an observable direct dependency of Youtube regarding the boundary resources that Epic provides. It is, therefore, explicitly not a platform complementor in the true sense of platform literature.

However, one might describe Youtube Gaming as a "complementing platform" to the Epic Game Store. As an independent platform, Youtube offers complementary services to the distribution network. It offers players, as well as game manufacturers, extended possibilities to engage with Epic's products, place advertisements or create additional content. As previously discussed, such a "complementing platform" relationship is also noticeable in the reverse direction.

Brandenburger and Nalebuff (1996) might provide a definition which renders a better fit for the presented scenario. They describe a complementor generally "as the developer of a complementary product" where products are complements if higher sales of one raise buying interest for the other.

While such a relationship itself appears plausible in the case of Youtube and Epic, there even is statistical evidence for such correlation (see Sherwin, 2019). We, therefore, adopt this definition comparable to earlier publications in the field of platforms. (see Gawer & Cusumano, 2014b)

An equally interesting as rare contribution to the relationship between two complementary platforms is provided by Eisenmann, Parker, & Van Alstyne (2011). They investigated the dynamics in the case of market entry of an existing platform into the market segment of another established complementary platform. Based on theoretical considerations, the prospects of success were evaluated, and it was concluded that "an entrant that bundles a complementary platform is most likely to succeed when the platforms' users overlap significantly". Although thematically different from the focus of our work, the paper nevertheless highlights the significance of the relationship between the two platforms for determining success factors. This paper also aims to make a valuable contribution towards the understanding and analysis of market dynamics between complementary platforms.

Research design

Having discussed the potential relationships between two platforms, the following section explores how these relationships can be analyzed in more detail. In order to answer this research question, the present study relies on quantitative methods on the basis of a case study and thus on an inductive research approach. Although the identification of potential influencing factors is theoretically motivated and therefore deductive in nature, the inductive elements of this study predominate. In order to gain the first scientific insights into this relationship of platforms, we will test theory derived factors in the setting of user interaction. We seek to investigate in a broad sense whether factors that are present in other areas of the platform interaction also influence the user interaction. Two factors, in particular, were important to us in selecting our cases for this study.

On the one hand, the observability via suitable automatable processes was an essential concern. Only then it would be possible to record the entire user behaviour over a longer period of time and to collect statistically valid data. In this context, platforms can be seen as a particularly valuable research target due to the accessibility of boundary resources. However, not every platform provides its users' data with the same degree of openness and quality. It was, therefore, necessary to find two complementary platforms that both willingly share their data at the same time.

On the other hand, the development of the entire industry and the overall ecosystem is a key benchmark for our selection of the case platforms. To ensure a constant and stable observation, the services and business models of the observed platforms should already be matured and not be subject to frequent changes. Only in this case, it is possible to evaluate the effects of user interactions without having to consider the side effects of external factors such as modifications of the platform itself or other ecosystem actors. It was, therefore, important to select an industry that has reached a developed stage rather than being in the process of being created or evolving. As already discussed in the theory section, in these industries, the uncertainty among the participants is also reduced, resulting in a more robust observed model.

An industry that meets these requirements can be found in the video game sector with the two platforms Twitch and Steam. Both platforms offer enough openness in their observable data, while the industry is well developed through decades of successful operation.

In the following, we want to present the overall video games industry as well as the concrete cases of Twitch and Steam in sufficient detail and place them in the context of platforms.

Case Description

Video Game Industry

In this section, we will introduce and describe the video game industry as a whole, which is the industry that our case companies operate within. Understanding the totality of circumstances within which our digital platform-cases of this paper exist, is crucial to interpreting the findings of this paper.

The video game industry has been on the rise for a long time, assisted by the technological evolution of both platforms, gaming equipment and additional periphery technology. In 2012 according to Statista, the video game market was worth approximately $52.8 billion, while in 2019, it had risen to $123.52 billion (Gough, 2018). As of August 2016, Statista research shows that people preferred gaming through mobile devices in the global market, followed by PC, social/online, and lastly on consoles (Statista, 2016). The gaming ecosystem is fairly different from other industries. However, the following graph from Hackernoon.com maps the video game industry ecosystem quite well. (Hackernoon, 2020)

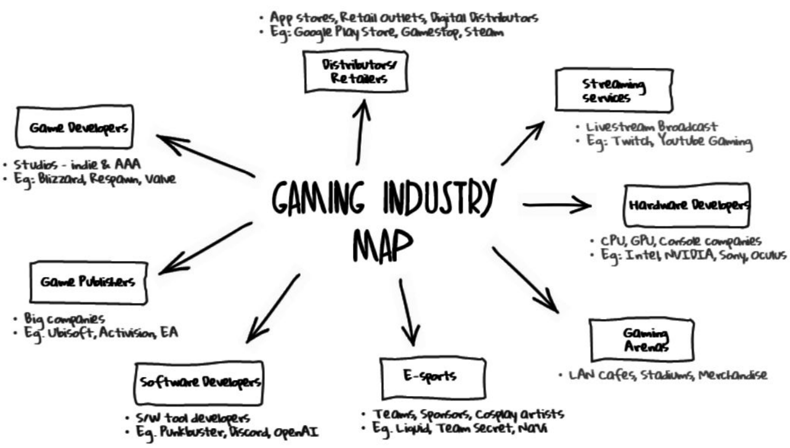

Figure 04. Map of video game industry landscape (Hackernoon, 2020)

The video game industry and the publishing of a game follows a certain timeline until release. It all starts with hardware developers developing the technology required to run the specific game. As an example, the hardware developers who invented the virtual reality system opened up for an entirely new type of games. The game developers then have the opportunity to develop games that run on this hardware, be it PC, PlayStation, Xbox, Wii, or mobile. The developers will often cooperate with the publishers, or try independently (indie) to release a game on digital distribution platforms. Once the game is released, other stakeholders are then engaged, such as streaming services, e-sports, gaming arenas and software developers. These external stakeholders all add additional value to the core service, which is the game itself. Streaming services such as Twitch and Youtube gaming allow users to share their gaming experiences with others live. Youtube itself, however, differs from Youtube Gaming and Twitch by allowing for upload of pre-recorded game-related content to be uploaded to the platform. They are incentivized by these streaming platforms to upload content consistently, as they make money through how much advertising they can attract and show to users who view their content. Software developers often develop periphery products to the games and are close to the definition of platform complementors in the platform literature (Wessel, Thies, Benlian, et al., 2017). These are companies such as Discord, who offer gamers a platform to communicate on through either voice or text while playing games. At last for the most popular of games, there are e-sports, which is a group consisting of sponsors, tournament organizers, as well as the most elite of players attempting to compete in a similar way to traditional sports for a cash prize. E-sports, in itself, has had an explosive development in the last decade. When combining occasional viewers and frequent viewers Statista (Statista, 2020) says that e-sports had a total viewership of 134 million viewers in 2012, compared to 454 million in 2019.

Market research

The way customers spend money in the video game industry has developed extensively over the last decade, as companies are attempting to figure out the most profitable and attractive business model for their games. These business models differ from platform to platform, and from game type to game type. Globally we have seen a switch from money spent in the packaging market, meaning games purchased in full in retail, to money spent on DLC (downloadable content).

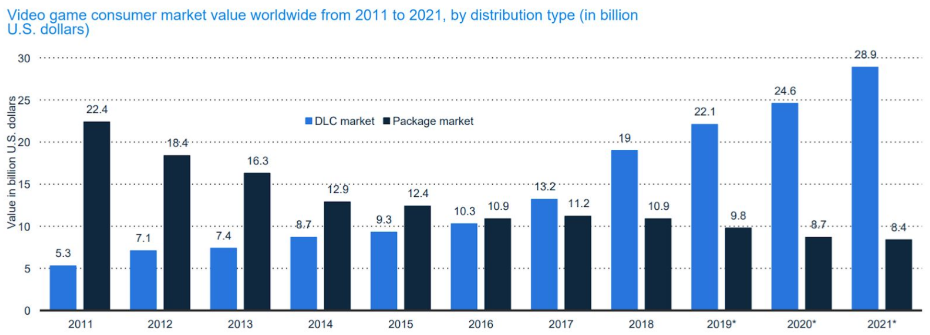

Figure 05. Video game consumer market value (Statista, 2018b)

DLC includes items such as modifications and transformations to traditional cosmetic objects within the game, commonly referred to as "skins". DLC at the source, Statista, also includes games that are digitally distributed. DLC is the main source of revenue within games following either the freemium model or the free to play business models, as they generate no revenue from sales. The global video game market has increased not only its generated revenue but also its pool of consumers (Statista, 2018b). The number of active video game gamers has increased from 1,8 billion in 2014 to almost 2.1 billion in 2016, an increase of about 16%. In the same period of time, the industry value increased from $71,25 billion to $93,29 billion, an increase of almost 31%. These data points show that the video game industry is growing not only in consumers but also in the amount of money spent per consumer, making the industry very lucrative. Statista further projects that in 2020 the market will reach $131,23 billion in value, and with a total of 2,6 billion users. This growth is again an increase of about 40,6% in value, and 28,5% growth in the number of active video gamers in the same period of time, showing that the trend of market value increasing more than the increase in consumers is consistent over a more extended time period. Most of these active video gamers reside in the Asia Pacific, having around 1,2 billion active video gamers, equalling to the Asia Pacific making up over 50% of the market in the number of active gamers.

External factors to the video game industry

There is a range of economic laws and hardware upgrades that causes the video game industry to grow continuously. Personal computers follow Moore's Law, which states that the number of transistors on an affordable CPU (Central Processing Unit) will double every 18 months (Cumming et al., 2014). Moore's Law gives game developers an increasing opportunity to create games that are very demanding on the hardware. Games with exceptional graphics or latency requirements are becoming continuously better, as the hardware around the PC improves, including the average speed of the internet connection owned by the average consumer. New technology also keeps emerging, as technology is allowing gamers to experience games in more realistic and immersive ways, such as virtual reality.

We will now present an overview of the digital platform companies used as the cases for this paper: Steam and Twitch.

Steam (Valve)

Historical View

Valve, the company behind the gaming platform Steam, is one of the biggest, possibly the most significant, company within PC gaming. Steam is the largest game distribution platform on the market, with around 75% of the market space in 2013 (Edwards, 2013). Their total sales in 2017 were around $4,3b USD, equalling around 18% of global sales for PC games (Bailey, 2018). That same year, developers released 7600 games on Steam, equal to a staggering ~21 games released per day. In short, there is no talking about game distribution platforms or PC gaming without mentioning Steam. Valve released Steam in 2003, and it was one of the earliest companies to leverage network economics within the gaming market by releasing a multi-sided platform, much earlier than Playstation (Sony) in 2006 (Sony, 2020). Since then, many other distributors have realized the potential and appeal of digital platforms, and have themselves launched their games as part of an internally owned digital platforms. Some examples of these companies are Blizzard with their "Battle.net" platform (Fahey, 2009) in 2009 and Epic games with their storefront launching in 2018/2019 (Frank, 2018). Epic games actually started out selling their games on Steam, but were unhappy with Steam's 30% cut of their revenue from sales (Frank, 2018).

Valve had their first significant success after the launch of the game "Half-Life" which was released as early as 1998, winning over 50 "game of the year" awards (Valve, 2020b). Besides launching Steam, Valve is also responsible for some of the most prominent and famous titles in PC gaming such as the Counter-Strike (CS) Franchise, which is one of the most played games in the world today and has survived for more than two decades. It is also one of the biggest e-sports in the world measured by viewers and price-pool (Hitt, 2019). Besides Counter-Strike, Valve also acquired the intellectual property for the modification of the Blizzard game Warcraft III, known as Defense of the Ancients (DOTA) (Onyett, 2011). Dota 2 is today the 2nd highest price-pool in all of the e-sports with a prize pool of around $46,7 million in 2019 (Hitt, 2019). If you've been playing PC games since the 2000s, you have undoubtedly heard of some of Valve's other titles, such as Left 4 Dead, Portal, Half-Life 2, and Team Fortress 2 (Valve, 2020b). Undoubtedly, Valve has had extensive success within the gaming market and should be seen as a titan in the industry. Led by CEO Gabe Newell, Valve continues to develop games, hardware and development on their platform Steam.

Actors on the platform

Steam is a digital multi-sided platform which means it connects two or more actors that would be hard or difficult to connect in the absence of the platform. In this section, we will cover the actors on Steam, as it is crucial to understand the platform, and thus our findings. There are two actors on Steam that make it function as a multi-sided platform: Players and game developers.

Players

The first actor on Steam is PC gamers. Within this project, and in the case of Steam, their customers are only PC gamers, as it is the only hardware compatible with Steam as a software. The PC gamers will be referred to as "players" for the remainder of this paper. While Twitch also has streams from Playstation and Xbox, Steam only has gamers who play on PC. Most of these players, 95,93%, are using Steam on a windows system, and the last 4,07% on a MAC operating system (Statista, 2018a).

Steam's relationship with the gamers is similar to that of a normal relationship between business and customer. Steam must maintain the gamers' interest in their platform by making sure that the games the players are interested in are available on the platform, as well as supply competitive infrastructure on the platform to make gaming as enjoyable as possible for the gamer. Maintaining this relationship includes giving the gamers access to communication through text messages and the ability to add friends and interact with friends through the platform itself. Steam has to maintain this relationship and their attractiveness to gamers to deliver on their value proposition. As with any other multi-sided digital platform, Steam is dependent on a high number of both supply and demand to interact on the platform to attract the other side of their multi-sided digital platform.

We will spend most of this paper testing assumptions and hypotheses about the demand side of Steam (players), as we operate from the assumption that it is the players who facilitate the cross-platform interaction with our other case: Twitch. The reason we chose to focus on the players in this paper, is that we are interested in the relationship between these two platforms, which will mainly be defined by its active participants. While we cannot say that one side of a multi-sided platform is more important than the other, we can say, however, that the users are the ones who are active on the platforms. The developers are much more passive on the day to day engagement on the platform, but the gamers will be playing different games, and give a much better view of the relationship between the two case platforms. We are in this project interested in researching the consumer behaviour on complementing platforms using the two selected platforms as a case study, and thus it only makes sense for us to focus on the active participants on both platforms.

Game developers

The second actor is the game developers, who supply Steam with a continuous supply of games, against having the ability and opportunity to reach millions of PC gamers through Steams platform. Both the developers and Steam are dependent on each other for continued business in most cases. However, in some cases, the game developers have enough bargaining power to gain independence from Steam, like the one we saw with Epic Games, who ended up launching their own platform for their game Fortnite to avoid Steam's 30% revenue cut (Frank, 2018). Valve is itself part of this segment of actors, as they themselves supply some of the most popular games to Steam, as previously mentioned.

Steam's relationship with the game developers is different from their relationship with the players. Steam must make sure to stay attractive as a distribution platform by maintaining a high number of users to purchase the developers' games. Besides this, Steam must entice game developers with an attractive economic contract that gives Steam sustainable revenue from the games but also incentivizes developers to use their platform to distribute games continuously. When Steve Jobs launched the app store, he famously pitched developers this idea of pure simplicity with releasing apps to iPhones. No fees, full control of the price point, and you are paid your 70% share of the revenue monthly (gamingandtechnology, 2008). This pitch is very similar to the approach Steam takes with their game developers. There could be hints, however, that the game developers are not happy with the current revenue split, as, in GDCs latest survey of nearly 4000 developers, only 6% thought Steam could justify their 30% cut (GDC, 2019). The previously mentioned Epic Games Store takes only 12% and is showing that the competition is increasing in the market. Steam may have to consider improving their relationship with their game developers, by in the future considering taking a smaller cut (Kuchera, 2019).

Business Model

Steam made $4.3 billion in-game sales revenue in 2017, with 63% of that coming from North America and Europe. They have been growing in new users every year. However, it looks as if the graph may be plateauing, as can be seen underneath.

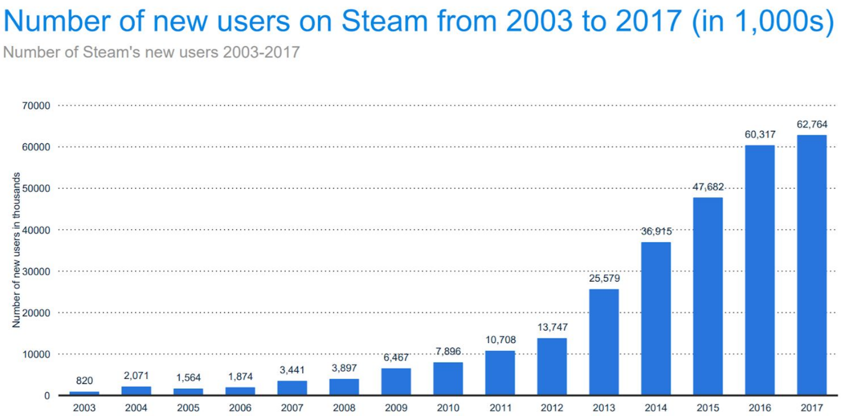

Figure 06. Steam user development (Statista, 2018a)

As previously described in the section about the video game industry, Steam generates revenue either through DLC or digital purchases. Either Steam generates revenue through the user purchasing the game itself to have indefinitely, or Steam generates revenue through free-to-play games, where there is downloadable content to be purchased through the app, such as cosmetics or resources. Steam takes a 30% cut of the revenue made by the developer. Steam also has a marketplace for different games, where players can sell in-game items for real money, where Steam takes a cut of the final price (Valve, 2020a).

Value Proposition

Steam's value proposition is that it is the most optimal place to go for digital purchasing and distribution of games, while simultaneously offering storage of games, game data, and a platform for community activities. Steam, in short, offers everything the user could want for PC gaming, including games for niche segments such as virtual reality.

For game developers, the value proposition is simply that the size of Steam's user base makes the revenue cut worth it for the game developer to distribute their platform on Steam. If this changes in the future, Steam may have to consider lowering the revenue cut in order to stay competitive.

Reasons for case selection

The reason why we have chosen Steam for this project as our platform of research is that it is the best option given all the criteria possible. Steam is the biggest platform within PC gaming, which is the vast majority of games streamed on Twitch from our data. Secondly, Steam is very extensive and broad in its game supply, meaning that it supplies a very large quantity of games, including games not developed by Valve. Battle.net as an example only supplies games that Activision Blizzard developed and published, thus limiting the research significantly. Third, Steam provides through its SteamSpy API and data distribution, an effortless way of acquiring data without needing permission or authentication by the company itself. All in all, there is no game distribution platform more suitable for this case study than Steam.

Twitch (Amazon)

Historical View

Twitch.tv is a live-streaming platform specifically aimed at gaming and e-sports (Geeter, 2019). Twitch launched in 2011, but before its rise to fame as a gaming live-streaming platform, it was known as Justin.tv, a website dedicated to following the life of one of the founders, Justin Khan (Geeter, 2019). The platform thus started as a reality TV platform, but when users started using the platform to broadcast their gaming experiences, the platform saw an opportunity for a pivot, and launched Twitch in 2011, solely focused on the live broadcasting of video game content by users.

The platform very quickly grew in popularity, being the first mover in a market where there was a high uncapitalized demand. This growth in popularity led to Amazon buying Twitch, for a staggering $970m in 2014 (Gittleson, 2014). Jeff Bezos, Amazon's chief executive officer, stated about the purchase "Broadcasting and watching gameplay is a global phenomenon and Twitch has built a platform that brings together tens of millions of people who watch billions of minutes of games each month," (Gittleson, 2014).

Twitch is essentially operating as a multi-sided platform connecting users who wish to stream their gaming content live and users who wish to watch live-streamed gaming content. Viewers can view Twitch from any internet browser or smartphone, and Twitch has millions of viewers every month.

On the streamer side, they allow the streaming of games through many different platforms. Users first and foremost stream their content from their PC. Besides this, Twitch also allows for streaming consoles such as Playstation and Xbox through integrated services on the platforms themselves (Andronico, 2019; IGN, 2016). It is also possible to stream Wii content directly to Twitch, however only through the use of a capture card, and not through any integrated service from Twitch itself. It is not possible to stream smartphone content directly to Twitch, which might be something they add in the future, given that mobile gaming makes up half of the gaming market as previously mentioned.

Twitch also channels some of their content through different APIs, that allow companies like Discord to show that a user on Discord is streaming content on Twitch.

Lastly, Twitch hosts a biannual convention called Twitchcon (Twitch, 2020a). Twitch con is according to Twitch themselves an event where "Everyone is invited to meet streamers, play games, watch esports, hang out with friends, grab new merch, and so much more" (Twitch, 2020a).

Actors on the platform

Viewers

The first actor on Twitch is the viewers, of whom there are millions. The average concurrent viewers for 2019 was around 1,3 million people (Leftronic, 2019). The average viewer spends 95 minutes daily on the platform, and Twitch receives more than 15 million unique daily visitors, which adds up to Twitch being the 35th most popular website on the internet as of the end of November 2019 (Leftronic, 2019).

Twitch's relationship with their viewers is that of a traditional business to customer relationship. Twitch wants to continue to offer great products or services, and wants to incentivize customers to spend as much money and time on the platform. Twitch achieves this by making sure their platform is easy to use, and that the platform has the best content out of its competitors on the market.

Streamers

Following the viewers, the next actor on Twitch we will describe are the streamers. The streamers and viewers interact on the Twitch by representing supply and demand of the digital multi-sided platform. Streamers supply streamed content for the viewers to watch on the platform by allowing the viewers to tune into their stream. Viewers can find streams to watch by filtering the main website by the games they want to watch, and they can follow streamers whose content they enjoy.

The "streamer" term encompasses any individual who chooses to broadcast their games on the Twitch.tv platform for others to watch. On average, there are around 3.7 million unique Twitch broadcasters per month, and in May of 2018, there were 27.000 Twitch partner channels (Smith, 52 C.E.). To become a Twitch affiliate, a streamer has to have streamed eight hours in the last 30 days, stream at least seven days in the last 30 days, receive an average of three viewers per stream, and grow their stream to more than 50 followers. Fulfilling these criteria will automatically get the streamer invited to the Twitch affiliate program.

Once a Twitch affiliate, a streamer's viewers will have the opportunity to "subscribe" to the Twitch streamer. A subscription costs $4.99 a month. Initially, the streamer and Twitch split this revenue 50/50, but for the more prominent streamers with the bigger audiences and higher number of subscribers, they earn a more substantial portion of their subscription income themselves (Consulting.com, n.d.).

Twitch affiliates can further their careers as streamers by applying to become a Twitch partner. Twitch will grant the status of partner based on the streamer's content, average concurrent viewership, stream frequency, and schedule. If Twitch finds that the streamer meets the demands of a partner, they will offer the partnership to them.

The main way streamers make money is through subscriptions by viewers. The subscription part of a streamers revenue streams is very consistent income, giving financial security to streamers. With a steady income every month, the job of full-time streamer appears more feasible for the streamers if they can attract a large enough audience consistently. Another way of earning money is through donations of either "bits" or real money to the streamer. Bits is Twitch's virtual currency, which viewers can buy and donate to the streamer. Twitch pays the streamer 1c per bit donated to them, while the current price for 25.000 bits is $308, ranging Twitch's cut of the money anywhere from 28% for the users who buy the $1,40 pack of 100 bits to around 19% for the users who buy the 25.000 bits. This currency facilitates the ability for viewers to put in smaller amounts of money and not be ripped by the transaction fee of a service like PayPal. If the users do not want to donate bits, they are also often welcome to donate real money to the streamer's Patreon or Paypal. Donations are often rewarded with attention from the streamer, and many streamers allow for the donation message to be displayed directly on the stream instead of in the chat function adjacent to the stream, mostly to encourage more donating (Influencer Marketing Hub, 2018). Streamers can also earn money through advertising, which again is reserved for Twitch partners only. Twitch pays streamers per view of the advertising that is automatically displayed on the stream when a viewer enters. They can also be paid through affiliate deals or sponsorships by specific companies. However, this part is not facilitated by Twitch directly, but also not opposed by the company (Influencer Marketing Hub, 2018).

Twitch's relationship with their streamers is similar to the relationship between a supplier and the business. Both sides have bargaining power, as Twitch can shut down the streamer's channel should they be uncooperative, but the streamer can also choose to go to a different platform. The bargaining power of both sides depends on the size of the streamer's audience, and particularly the number of viewers that would follow the streamer to a different platform.

Business model

Twitch's business model is a mix of several different business models for digital platforms. Twitch has several income streams through its platform. The first source of revenue we will describe is their revenue generated through the Twitch Prime subscriptions. Twitch Prime is a subscription service that lets customers earn in-game rewards for specific games, and play other games for free, rotating each month (Twitch, 2020b). It also allows the user one free Twitch subscription to a stream, worth about $4.99. The subscription to Twitch Prime costs $5.99 a month after a seven-day free trial, and the user on top of the benefits on Twitch also receives benefits on their Amazon account. In addition to this, Amazon Prime customers can subscribe to Twitch Prime for free.

The second revenue stream for Twitch is their subscription service called Twitch Turbo. Twitch Turbo is essentially purchasing the premium Twitch experience, giving the user a smoother viewing experience by removing advertisements across Twitch, expanding the user's access to chat and emoticon colours and types, as well as giving the user priority on customer support (Twitch, 2020c).

Further than that, Twitch generates income through the subscriptions viewers make to individual streamers, costing the viewer $4.99 per month. Twitch takes half of this revenue (Fortney, 2018), generating them about $2.50 per subscriber. The top 10 streamers currently have around 265.000 subscribers when their gifted subscriptions, prime subscriptions, and paid subscriptions are combined for each streamer.

Twitch also generates revenue like a traditional digital platform running the freemium business model through advertising. Twitch does not disclose its revenue from advertisements directly, but

Newzoo predicts the video game market to top $906 million in revenue with advertisements and sponsorships making up the most of that revenue (Fortney, 2018; Takahashi, 2018).

As previously described, Twitch also offers a service called Bits. Bits are Twitch's virtual currency, and as described in the description of the streamers, Twitch retains between 19% and 28% of the money spent by users purchasing bits, while the rest goes to the streamer to whom the user donated.

Value Proposition for Viewers

The value proposition of Twitch should explain to us why the viewers are deciding to choose Twitch over Youtube gaming, Mixer, or Facebook. Twitch's value proposition is directly related to Twitch holding around 73% of the market value (Valentine, 2019), as the value proposition helps us understand why the viewers make the choice to go to Twitch and watch streams.

Twitch's value proposition centres around being a digital platform, which means creating an easy and stable website that facilitates the interaction between different user groups. Twitch adds value through the ease of use for streamers, and the simplicity for viewers. Viewers have to only go to one website to watch their favourite games or their favourite streamers. Twitch is the market leader, which means they through network effects and winner-takes-all dynamics attract the most viewers and the best (and most) streamers. In simple terms, the customers should choose Twitch over the competing platforms for game-related content, because Twitch offers the best content at the time of the users' desire to watch video game live-streams. They on top of this core value proposition offer good infrastructure on their website, with easy to use ways for viewers to support their streamers.

Value Proposition for Streamers

As with Twitch's value proposition for viewers, Twitch's main value proposition to streamers is related to their size. Twitch is simply the most profitable platform to be broadcasting on as a streamer, because of the size of the audience frequenting Twitch. Twitch offers a transparent business model and career path for streamers which incentivizes streamers to broadcast on the platform.

Streamers are a crucial asset for Twitch, and thus the relationship between Twitch and their streamers is most accurately compared to the relationship between a supplier and a business. The most optimal situation is a continued cooperative relationship where both partners are happy and depend on each other. However, if streamers become uncooperative, Twitch has the power to shut their stream down, and if the streamer finds that Twitch is unreasonable, they can move their stream to a competing platform. The bargaining power of both parties depends on the size of the streamer's audience, and in particular the percentage of that audience that would follow the streamer to another platform.

Twitch needs to continuously make sure that streamers are satisfied with the business arrangements and conditions of streaming on their platforms, and that competing platforms are not offering better conditions for streamers. Twitch has a major advantage they naturally gain from the size of their audience, which mainly comes from them being the first mover in the market. Setting aside the size of the platform, competitors like Mixer, Youtube Gaming, and Facebook Gaming, are trying to edge in on Twitch's market leader position by differentiating themselves from Twitch and trying to lure the most popular streamers to their platform. Thus the relationship between Twitch and their streamers is an essential part of the continued business for Twitch.

Reasons for case selection

In our description of Steam, we described why we chose Steam as the game distribution platform for this project. We will now also introduce why we chose Twitch as the platform of research. When we are talking about live streaming of video game content, there are only four platforms that would make up the market: Twitch (Amazon, 73%), Youtube Gaming (Google/Alphabet, 21%), Mixer (Microsoft, 3%), and Facebook Gaming (Facebook, 3%) (Valentine, 2019). While attempting to gather data from all the platforms was considered, we ultimately decided that it was outside the scope of this research. We realized that such a task would require us to store much more data than we had the capability to, and so we wanted to choose a platform that was singularly representative for the market. Twitch being the market leader, we felt was the obvious choice. Twitch not only represents 73% of the market of live stream gaming in 2019 (Valentine, 2019) by hours watched, it has also been around for the longest and is, therefore, the most established. In addition to this, they had an accessible API, that allowed us to efficiently scrape the service for data, allowing us to stay within the scope of the project for the data collection phase. In short, Twitch is 73% of the market, the most established and also the easiest to gather data from. This makes Twitch the optimal choice for the case study out of the platform and its competitors.

Selected Case and Complementary Types

As already discussed in the theory section, the interdependence between watching live streams and playing a game on one's own is a strong indication of the increased utility value of shared usage of both platforms Twitch and Steam. We, therefore, assume in the following that the platforms Twitch and Steam are complementary to each other.

A complementary nature between two goods may, nevertheless, take different shapes and characteristics. Thus, we would like to analyze different types of complementary goods and put them into context with regard to digital platforms and services as well as the specific case of Twitch and Steam.

One of the major contributions to the nature of complementary relationships in recent years can be found in the paper "One-Way Essential Complements" by Chen and Nalebuff published in the year 2006. The authors fueled the debate on the distribution of profits between complementary product producers by carefully reviewing a very specific two-goods scenario. This scenario is best described as "the case where one good (A) is essential, and the other good (B) is optional. A consumer can enjoy A without B, but not B without A".

To illustrate a constellation that follows these rules, the authors provide the example of the Windows Operating System (A) and a media player (B) that can be installed on it. While the media player is highly dependent on the operating system, the operating system can also generate benefits for the user without the media player. Transferred to today's platform systems and software, one could apply the same definition to the Apple App Store platform for iPhones (A) and an app developed for this platform as market participant B.

But it is not only the prominent example of Apple that demonstrates the growing importance of "one-way essential complements". Almost all digital platforms, such as PayPal, Facebook, Twitter, Shopify, or Amazon, also created ecosystems through highly dependent third party developers to further evolve and innovate. This open co-creation approach is both a major opportunity and at the same time challenge in the context of platforms, attracting a considerable amount of scientific interest. (See Gawer & Cusumano, 2014a; Nalebuff et al., 1996; Rietveld et al., 2019)

All the more, it becomes evident that for today's analysis and evaluation of platforms, the research by Chen and Nalebuff proves to be profoundly relevant.

When applied to the present case of Twitch and Steam, a similar relationship can be observed between the games and live streams on a superficial level. One could easily understand the game as an essential good because, without it, the live stream would not be possible. However, this hastily made assumption involves a number of misconceptions. On the one hand, only for the streamers themselves, the game is an essential good to participate on Twitch and start live streams. For the audience, which is the main focus of our observation, the game is not a prerequisite to watching others play on Twitch. Secondly, a comparison between the platforms in terms of their complementary character is not limited to the games or live streams offered on them. The services offered by the platforms go far beyond the games and live streams provided on them, as their value for the user is generated in particular by other network effects. Limiting the assessment of the complementary solely to the products offered on the platform does not adequately reflect the complex services offered by Twitch or Steam.

Neither the service of distribution and organization of games offered by Steam nor the support generated by Twitch through the hosting of live streams and the creation of a vibrant community is dependent on the respective other platform providers.

Consequently, the examined companies also do not correspond to the two-way dependent complements examined by Farrell and Katz (2003). In their research, the two economists examine the consequences of innovation and competition for components in the computer industry. The key limitation here is that the "Components A and B are valuable only when used together" (p. 413). This form of dependency occurs, for example, in the interaction of hardware and software products.

Schilling (2003) further discusses the possibility to realize lock-in effects in the video game industry by promoting this kind of complementary products. With the case of Microsoft and the Xbox platform, the author illustrates that this kind of complementary relationship is also highly relevant in the platform as well as the games and software industry especially. Nevertheless, even this market scenario is not comparable to that of Twitch and Steam.

The market situation examined in this study is, therefore, not a complementary configuration that has been specifically recognized or developed in the Platform literature yet. It can, therefore, be perceived as a rather generic constellation of complementary services in which neither is essential for the other.

To define this generic constellation in more detail, we could discuss to what extent Twitch acts as a complementary partner for Steam and vice versa. For this purpose, it would be particularly interesting to assess the mutual utility of the services that the platform provides for each other.

The Steam platform, on the one hand, enables users to buy and organize video games. Thereby it enables users to translate the live streams seen on Twitch into gaming experiences they can play themselves. There is, however, no in-depth integration of live streams or community features within the Steam platform. As the de facto market leader in the field of game distribution and the resulting lack of demand for innovation, the provider has little incentive to integrate features of other platforms. The coupling of the platforms in this direction may, therefore, be considered rather moderate.

On the other hand, the coupling of Twitch to Steam manifests itself more intense. As a media platform, the service enables a deeper level of interaction both on a technical as well as user level. For example, on a technical note, Twitch uses the data provided by Steam to determine which game-specific users are currently playing. In addition, through its very community-centric approach, Twitch offers many social interaction opportunities with the games distributed on Steam. An example of such interaction can be observed in the community matches that the Twitch-streamers host together with their viewers on Steam. In these tournaments, prominent Twitch streamers invite their audience to join large player vs player matches. This intense interaction with their followers not only creates positive bonding effects between streamers and their audience but also shows the deep interdependence of the platform Twitch on Steam.

Even if the dependencies of the platforms among themselves are not symmetrical, they still demonstrate strong tendencies of a complementary relationship and are therefore considered as complementors hereafter.

Hypothesis Development

The previous sections addressed empirical findings and theoretical considerations regarding the relationship between platforms, complementary goods, and other fundamental platform effects such as network externalities or modularization. As this thesis aims to explore factors that significantly influence user interaction between two complementary platforms, it is necessary to define user interaction between the case companies first. After that, we need to identify potential and influential factors, which can then be quantitatively verified in a third stage.

Thus, the discussed findings and existing platform research shall now be applied to develop testable hypotheses considering different theoretically motivated factors in combination with the chosen cases of Twitch and Steam.

Players and viewers as user-platform interaction

To first define the dependent variable of our analysis, we have to understand how the user-platform interaction can be modelled in the specific case of Twitch and Steam.

Looking first at the platform Steam, it becomes evident that it offers many and above all diverse opportunities for users to interact with the platform. The possibilities range from buying games, chatting with other participants, commenting on games and discussions to collecting achievements. However, all these interaction patterns can be considered secondary in comparison to the actual playing of games as the primary interaction with the platform. The excellent operationalizability, as well as observability through actual player numbers per game on the platform, is another advantage of using this interaction metric as a dependent variable for the game distribution network.

On the Twitch side, the interaction opportunities for users can be similarly diverse. Commenting and linking streams, discussions in forums or buying discord credits are just a few of many small components and building blocks of a complex interaction pattern. To reduce this complexity, we also limit our observation in this case to the primary metrics of the number of viewers in streams on a game basis. The operationalizability and observability of the variable remain the primary objectives in this area as well.

One possible objection to this approach could be that interaction with the game, either by playing on Steam or by watching Twitch, is not an interaction with the platform itself, but with the underlying game application. However, we are convinced that such a separation between platform and game is not apparent due to the strong connection between the two services. The game cannot be observed outside its platform provided ecosystem. The game offered on the two platforms represents the observable link between Twitch and Stream and thereby makes the different interaction with this product observable in the first place. It is therefore not even possible to evaluate the interaction between the platforms independently from the product offered - in this case, the game.

We, therefore, understand user interaction as player interaction and viewer interaction. In the following, we seek to identify factors explaining the variance across the platforms.

Life Cycle

As outlined briefly in the discussion of the study by Adner and Kapoor (2010), the concept of product life cycles in relation to innovation ecosystems plays an important role in platforms. Also, in the context of user interaction, the life cycle of a product or technology could have a significant impact on user behaviour.

The current literature fundamentally distinguishes three different dimensions with regard to the life cycle. These streams of literature try to cluster their research ambitions around the concepts of the *industry life cycle*, *technology life cycle* and *product life cycle*.

The *industry life cycle* thereby represents the most distanced perspective of the three concepts on economic market trends. Quite generally speaking, it focuses on the time and development that an entire industry undergoes between its creation, decline and disappearance. Oliver Williamson (1975) describes three phases of such a development in particular "an early exploratory stage, an intermediate development stage, and a mature stage. The first or early formative stage involves the supply of a new product of relatively primitive design, manufactured on comparatively unspecialized machinery, and marketed through a variety of exploratory techniques. [...] The second stage is the intermediate development state in which manufacturing techniques are more refined and market definition is sharpened, output grows rapidly in response to newly recognized applications and unsatisfied market demands. [...] The third stage is that of a mature industry. Management, manufacturing, and marketing techniques all reach a relatively advanced degree of refinement. Markets may continue to grow, but do so at a more regular and predictable rate".

The *technology life cycle* and *product life cycle* are two concepts that are often considered difficult to differentiate by the literature. (See Taylor & Taylor, 2012, p. 542)

While the technology life cycle usually measures the time span within an industry, "between a new technology's emergence and its decline", the product life cycle can be defined as "within a given technology, the time span between the emergence of a new product and its decline" (2010, p. 314). In the world of platforms, however, technology and product are strongly interlinked, as the example of Amazon Web Services illustrates. On the platform, products are marketed that represent the entire technology stack and therefore blur the lines between product and technology.

Although the various life cycle models examine differently granular entities from industry to product, all concepts describe a similar pattern represented by the phases of *market development*, *market growth*, *market maturity* and *market decline*. (See Levitt, 1965)

In the case of complementary platforms, these platforms often share a common life cycle of the product or technology developed on it. Amazon Web Services, for example, share the same life cycle with their complementary digital ocean in many related products like database technologies or container orchestration services. Another example could be mobility service platforms like Uber or Flixbus where both platforms offer different products but are, to some extent, sharing the technology as well as industry. We assume that the impact of the life cycle goes beyond the field of innovation between platforms and has a significant influence on the relationship between two platforms on the user side.

In the case of Twitch and Steam, the platforms offer different services on the same games. These games are in different stages of their specific product life cycles affecting the interaction of the users with the platform to varying degrees. The following example can illustrate this relationship. Both the game Grand Theft Auto IV, released in 2015, and the recently published game Football Manager 2020 are available to play on Steam and stream on Twitch. While Grand Theft Auto IV is likely to be at a late stage of its product life cycle and will not attract many new players, Football Manager 2020 is a recent game that has just entered the market. These games differ significantly in the way users perceive them, which also affects their interaction with the platform. For this study, the primary factor of interest lies in the different effect of the product life cycle of a game on Steam players and Twitch viewers. Does the attractiveness of a game develop in the same proportion on both platforms as the product life cycle progresses or does one platform enjoy advantages over the other?

We suspect that the attentiveness half-life period of a game on Twitch is significantly shorter than on Steam. On the one hand, we assume that streamers prefer newer games and therefore offer them for streaming, on the other hand, we expect that players with a long-standing affinity to a game rather play the game than watch it. These considerations lead to the following hypothesis.

H1: A progressing product life cycle decreases viewers more than players

> **H1a:** A progressing product life cycle decreases players
>
> **H1b:** A progressing product life cycle decreases viewers

Social interaction

The second key aspect that is extensively discussed and analyzed in the platform literature can be summarized under the concepts of network effects. These - among others - positive reinforcing effects associated with the utility of a platform can occur not only between two platforms but also for products displayed on them. In the case of Twitch and Steam, the games offered on the platforms may include so-called multiplayer capabilities. In these game modes, the players do not play on their own against the computer, but rather against or with each other connected via the internet. In this respect, games can also be described as mini platforms that allow the social interaction between the players to create utility for the overall game experience. Simple features such as a voice and text chat between the various participants can also contribute to this interaction, as well as complex user interactions within the game engine.

In order to avoid any confusion regarding the platform literature and already established concepts, we do not refer to the characteristic of multiplayer games as network effects but instead, use the term social interaction.

Applying this variable in the context of the cases Twitch and Steam, several implications are conceivable. We assume, and we expect that an increased social interaction on Twitch's side will lead to an increased entertainment value of the game and thus to increased viewership. We further expect that social interaction within a game will also bind more players to play the game. Beyond that, it is difficult to estimate which of these two effects will bind users to the specific platform with greater intensity.

From our own experience, we assume that the increased entertainment value of streams through multiplayer can have a greater impact on the viewers than the actual gameplay effects on the player side. We, therefore, assume the following relationship as the second hypothesis.

H2: Product social interaction increases viewers more than players

> **H2a:** Product social interaction increases players
>
> **H2b:** Product social interaction increases viewers

Access Barriers

An additional effect that is often discussed in the context of platform ecosystems and the management of innovation is the barriers to market entry for other participants. (Eisenmann et al., 2011, p. 17). A mostly empirically unexplored phenomenon regarding the relationship between two organizations, however, can be seen in the barriers that users experience when deciding to consume a product. Particularly in the context of platforms, where the consumption of a good does not imply its ownership, special rules apply to barriers of access.

These considerations are supported by the research paper "Burdens of Access" written by Hazée et al. (2017). Through qualitative interviews, the team identified various categories of customer-perceived barriers to access-based services. In principle, these hurdles can be divided into functional and psychological barriers. Whereas functional barriers actually involve the product itself with factors like complexity and usability, psychological barriers affect user behaviour through factors like brand fit or anticipated social status through the use of the service. From our distanced perspective of the ecosystem, we are only able to observe functional factors of the product. We, therefore, examine how the complexity of a game is perceived on the different platforms. Applied to the case of Twitch and Steam, an interesting dynamic might be observable. An access barrier on the side of Steam may, in fact, lead to an increase in user interaction on Twitch. Because the platforms in some ways offer substitutable ways of experiencing the product, access barriers as a whole could conceivably limit interaction on one platform, but increase it on the other.

Furthermore, we extend this metric to include a factor specific to the video game industry, which is the system requirement of a game. We expect that increased system requirements of a game represent a greater access barrier than more hardware independent games. An example of this phenomenon might occur in the virtual reality game sector, which developed in recent years. Typically VR games require not only more hardware components like virtual reality headsets but more powerful graphics cards and processors as well. This leads to the situation that a large number of casual gamers do not even get these games working on their desktop computers. Similarly, other technological advances such as ray tracing or game physics affect the requirements regarding computer hardware.

From marketing theory, furthermore, we can also derive the most obvious barrier for consumers - economic risk and financial loss. The expenses for participating in the service tie up financial resources that consumers cannot spend elsewhere, for example, on food or rent (See Ram & Sheth, 1989). This exposes them to an economic risk which consumers may perceive as restrictions or barriers on access.

From these considerations, we subdivide the concept of access barriers into three sub-constructs of complexity, economic risk and hardware requirements.

H3: Product access barriers increase viewers more than players

H3a: Complexity decreases players

H3b: Complexity increases viewers

H3c: Economic risk decreases players

H3d: Economic risk increases viewers

H3e: Hardware requirements decrease players

H3f: Hardware requirements increase viewers

Operationalization of the concepts

To verify the theoretical assumptions, it is necessary to quantify the constructs of the *Product life cycle* and *Social interaction* as well as *Access Barriers*. The scales used for this purpose are introduced and discussed in the following chapter.

Life Cycle

For the operationalization of the life cycle of the games, no trivial standard has been defined. In the case of Steam and Twitch, several factors come into question that could make the life cycle measurable.

On the one hand, one could consider whether the existence of a sequel game title as a binary variable could represent the estimation of a progressing product life cycle. Unfortunately, this metric has multiple shortcomings. At first, this measuring method can be perceived as relatively inaccurate because it basically only records the de facto discontinuation of the product. Secondly, when defining a sequel in the games industry, the question arises as to how comprehensive a new release of a game series must be in order to be considered a complete successor. There is a trend in the video game industry to offer many small additional products called Addons or DLCs after the main game is released. The bandwidth within these additional products is extremely wide and can span from single in-game items to very elaborate game packages. Therefore the border between DLC and sequel becomes blurred rendering the definition of a sequel game title unsuitable. A second trend in the video game industry which makes this measure seem debatable is the increasingly popular type of service games. These games are not provided with a normal release cycle of games but are designed to be played for a longer time period. The main game is not sold separately, but instead new content and participation in the game have to be purchased at regular intervals, so-called seasons. In these cases, there is no sequel at all as the game evolves in its original structure.

By contrast, the sales figures of a game could provide a useful measure for the advancing life cycle of a game. The disadvantage of this method of measurement lies in the fact that it is not sufficient to analyze the sales figures at the moment of the survey. Instead, historical data since the release of the game is needed to estimate the baseline of sales we can compare the current figures to. Unfortunately, this historical observation is not available to us, given that we only had one month of data recording.

Due to this constraint, we have adopted a simplified approach to life cycle assessment. We operationalize the life cycle of a game over the age of the game. In other words, the time that has passed since the game was released until the time of data recording. This seems particularly appropriate as games are not timeless products. Games follow a rather normal pattern as they become less attractive over time. Especially the rapid innovations in the field of computer graphics can quickly make a game look outdated over time. Furthermore, the age of a game can be reliably retrieved or calculated in an automated way.

The construct of the life cycle is thus determined in the following by the one-dimensional age of the game and thereby creates a continuous scale.

Social Interaction

In contrast to the life cycle of a product, social interaction in games is a relatively soft construct that depends largely on the perception of players. The operationalization of this factor is, therefore, relatively complex and multi-dimensional. From the outside perspective, however, this internal perception of the players is difficult to capture. One possibility arises, however, if you analyze user-generated data on the platforms, for example, user comments about games or streams, ratings of games and chat histories may be sources for such a user-centric view.

An automated way of retrieving such data in this context is provided on the Steam platform in the form of user-generated tags. These descriptive keywords associated with a game summarize the characteristics and features of this game in the shortest possible form in order to give potential customers a quick understanding of the entire product range. Through multiple tags per game, users can vote on the features and thus even display gradually to what extent a game contains multiplayer capabilities.

We, therefore, operationalize the construct of social interaction solely through the multiplayer capabilities of the games on a continuous scale.

Complexity

Estimating the complexity of a game is also related to the difficulty perceived by users. Therefore we follow the approach of evaluating user-generated tags analogous to social interaction. The tag of *difficult* is used for the purpose to also create a continuous scale of the complexity of a game.

Economic Risk

The economic risks of acquiring a game are relatively straightforward compared to other and larger consumer investments like cars or real estate. Indirect costs are low and are usually reflected in the electricity needed to play the game. Although some games offer non-transparent financing options, such as in-game currencies, and are therefore closely related to gambling, the economic risks are typically limited to the initial investment. The price of a game can vary greatly, depending on the pricing-model. While so-called free to play titles are even offered for no charge, there are also games on Steam that exceed an initial investment of 80 dollars. Exceptions are games like "Ascent Free-Roaming VR Experience" which cost 1000 dollar and are intended for VR enthusiasts.

We assume that the economic risk factor can be operationalized relatively well through the price of the game and likewise be included in the analysis as a continuous scale.

Hardware Requirements

The hardware requirement factor is probably the most complicated concept regarding operationalization within this project. On the one hand, the system requirements of games are not uniformly standardized, and on the other hand, many factors are involved in the performance of a system so that the measurement of such a requirement can rapidly become highly complex. To fully comprehend what this means, we need to dive more deeply into the mechanics of system requirements.

The first idea of operationalization would be to compare the requirements of the games on a purely numerical basis. For example, some games require a specific amount of RAM in the unit scale of gigabytes. This requirement could be read as if a game that consumes one gigabyte poses fewer limitations than a game that requires two gigabytes. But this is not the case.

Multiple factors besides just the amount of memory matter regarding the performance revealing the fundamental problem behind system requirements. Among other factors, the RAM revision and the frequency of the memory modules have an impact on perceived system speed. However, only a few games provide such detailed information, making it particularly hard to assess the specifics regarding their own system even as a well-informed user.

This confusion surrounding the performance characteristics of the individual components is not limited to computer memory. Also, the processor and graphics cards have similar performance indicators such as gigahertz and cache memory which only serve a purpose within their hardware generation. In addition, compatibility limitations between the various components and manufacturers may also lead to performance improvements or losses. Consequently, it cannot be recommended to analyze and compare the system requirements on a component level.

Another possibility which instead considers the overall system might analyze the price of the overall hardware. A total cost of the interoperable system does not consider different individual performance indicators of the components but rather a barrier represented by the overall monetary value. Nevertheless, also this approach bears severe disadvantages, especially in the highly volatile technology sector. Two factors, in particular, speak against it: The rapid technological decay with the corresponding price reduction of the components, as well as differences in price introduced through overall volatility of the market.

If we consider the memory produced by the manufacturer G-Skill as an example, this effect is clearly noticeable.

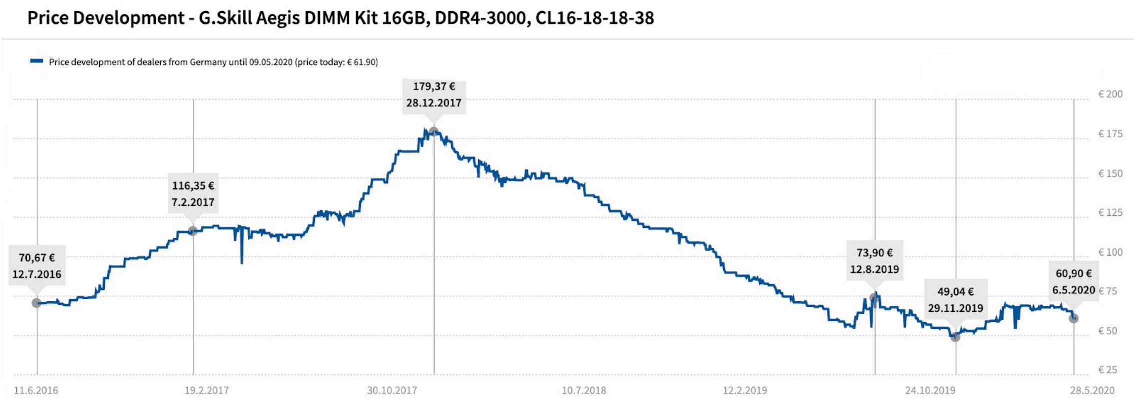

Figure 07. Price development example RAM (Heise Medien, 2020)

Figure 07 shows the retail price for a specific set of 16GB DDR4 memory over the last four years. It can be recognized clearly how volatile the market price has evolved over the last years. In the period from June 2016 to September 2017 alone, the price more than doubled, revealing a trend which reversed in the following years. Also, in the last year itself, considerable price fluctuations can be observed within just a few months. Between August and December 2019, for instance, the price fell from almost 74 € to 49 €, reflecting a loss of 33%.

Since similar pricing fluctuations can also be observed for the other components, the conversion of system requirements into price restrictions does not appear appropriate either.

A much more elaborate method to determine the genuine performance of computer systems is referred to as benchmarking. This method is not intended to prove the exact performance of a computer system, but to establish comparative results against other systems. Such a method is generally employed when testing components such as processors or graphics cards. So-called synthetic benchmarks can measure the theoretical performance of the components and put them in relation to each other (See Maxon Computer, 2020). The method does not only consider individual components but is also able to create reference values for entire computer systems. This facilitates the comparison of different systems in their individual configuration.

Comparing these results of the benchmarks with the requirements of the games, a so-called System Requirement Index can be calculated. This index value enables the comparison of the system requirements on a constant scale (See Game System Requirements, 2020).

Data collection

In order to test the established hypotheses within an appropriate framework, valid data on the various operationalized concepts must be obtained, stored and processed. We will describe in the following which challenges arose during the course of implementing the collection of the data set and how we addressed and solved each of these obstacles.

The fundamental objective here was to observe and record the variables defined in the previous chapter as accurately and with the highest possible resolution for the longest possible period.

Furthermore, the secondary goal was to record as many potential control variables as possible to enrich the primary data collection. Besides the pure quantity of variables, it was furthermore important to maintain their validity and reliability over the collection period.

At the same time, the given constraints such as storage space and query speed should not be exploited at the expense of the accuracy regarding our primary variables. In this dualism, it was especially important to achieve an appropriate balance between the amount of data collected and its relevance for this research.

Choice of collection approach and study design

First and foremost, it is essential to determine which type of study design is appropriate for the collection of the variables in question. Either an experimental design or a correlational study may be considered.

While in the former case variables will be manipulated by the experimental design and the execution of the study in order to measure the impact of these adjustments on other data, in correlation studies a phenomenon is observed in its natural context in order to identify possible relationships.

While experiments generally take place in a controlled environment and thus provide more precise insights concerning the correlations, it is often difficult to generalize these with regard to the underlying population.

This artificial character is less pronounced in classic correlational studies that focus on the observation of natural phenomena.

In the present case, however, it is not only this characteristic which speaks in favour of monitoring the state of the system as a whole. In particular, the manipulation of variables and subsequent observation in an experimental setting is unimaginable with the sheer size and global distribution pattern of the Steam and Twitch user base. The manual intervention and following monitoring of parameters with a small subset of users would only be feasible under extreme laboratory conditions, which would make the results and the generalization of these results questionable. Besides ethical questions regarding the manipulation of human behaviour, correlational studies also do not require complex experimental setup.

These arguments speak in favour of a data collection applying a traditional observational approach which maintains to be further supported by supplementary technological framework conditions. The research target platforms of Twitch and Steam provide extensive resource boundaries, so-called application programming interfaces, which are perfectly in line with the discussed and favoured approach. As a result, a substantial volume of the data to be collected is publicly provided by the platforms themselves as a primary source and therefore available directly to this study. The recording and enrichment of the supplied data remained the main task of the data collection regarding this thesis. In particular, the filtering, preparation and aggregation of the data happened to be a challenge which we address in the next section.

For the reasons discussed above, and in particular because of the unrestricted availability of primary data, we made a clear choice to pursue a correlational approach in order to acquire the relevant dataset.

Data retrieval and storage

In the following, we will show how the primary variables regarding the number of players and spectators were assessed and recorded. In addition, we will illustrate how the components of complexity, price and system requirements of a computer game, as well as social/multiplayer character of a title, have been captured together with rich metadata on both games and Twitch streams. The aim was not only to create a momentary snapshot of the gaming and streaming markets but also to capture all these variables in their variance and dynamics over time during the observation period.

Basic procedure

The broad undertaking of collecting the necessary data consists of a multitude of small sub-decisions and tasks which are far too numerous to be discussed herein sufficient detail. However, we would like to introduce and discuss the most important concepts and considerations on this page.

The primary objective of the data collection was to record as broad data as possible on players per match and spectators per stream. This means in the concrete case to include as many games as possible in the data collection.

Determining the maximum sample size

The initial approach could thus be to simply record all the games available on Twitch and Steam, including their player and viewer counts. However, by evaluating the raw numbers, it soon became evident that Steam alone holds over 30,000 games in their portfolio while Twitch offers as many if not more streams to follow. If one wanted to query and store all this information together with its metadata every 15 minutes for a month, one would need not only an immense number of requests and traffic but also a huge database with an immense amount of storage space. For this case, we anticipated and estimated a required storage capacity of over 200GB, which in our experience proved to be too large for the required system performance. As the size of the database increases, more complex and thus longer queries are required when adding more rows to it. Inserting data to such a large database of 200+ GB would have consumed too much time with the amount of input to support a maximum script runtime of shorter than 15 minutes.

As a logical consequence, it was necessary to limit the number of games requested on both the Twitch and Steam sides and to reduce them to a manageable scale.

In addition to the limitations imposed by the amount of storage capacity on the server itself, Steam and Twitch on the provider level also enforce guidelines for the usage of your boundary resources, defending and limiting their exploitation. The operational implementation of these guidelines manifests itself in two separate ways, a legal/organizational and a technical one. The first can best be described best as a "fair use" clause which allows the provider to restrict access in the event of inappropriate or excessive activity. The technical implementation called rate limiting restricts the number of requests to the different APIs over certain periods of time (Sturgeon, 2018). For example, Steam only allows 100,000 requests per day, which at first sight may seem like a rather liberal limitation. If this amount of requests is reached prior to the end of a 24-hour window, one has to wait the remaining period of time to be able to submit new requests. (Valve Corporation, 2010). Twitch is even more generous regarding its limitations by allowing 800 requests per minute and thus a total of 1,152,000 requests per day on their service.

However, examining these limitations in the specific case, reveals that the seemingly generous limitations quickly reach their limits in practical applications. Consequently, the following task involved selecting and extracting a representative sample from the population, which, in size, would fit the given framework conditions. The main objective of this process must be to include as many games as possible from the population while the selection does not endanger the overall sample's representativeness.

The maximum sample size with respect to the Steam API, which has the most restrictive constraints, amounts to about 1000 games. Querying the number of players for these games in quarterly intervals will result in a total of 96.000 queries per day, which is below the limit of 100.000 requests. The absolute maximum sample size can thus be determined to exactly 1041 games.

Selection of the sample

In order to select the representative subgroup, it is necessary to define the population that this sample should actually represent.

Since this study is fundamentally about the effects between platforms, exemplified by Steam and Twitch, the games are supposed to be available on both platforms. A game that is not present on one of the two platforms can thus not be included in this study in a meaningful way. Therefore, the population of all games are all the titles that are available on both Twitch and Steam.

The identification of such titles, being available on both platforms, turned out to be an unexpected challenge in itself.

With Steam it is possible to retrieve all products available for sale and thereby obtain the entire product catalogue. On the side of Twitch, however, only the approximately 2500 so-called "Top Games" can be queried, representing the current most viewed games on Twitch. This relatively small number can only represent a fraction of the publicly available games as all feature at least a single-digit number in viewers. Furthermore, the new Twitch API offers the possibility to request information about the game by providing the exact name of the title. A helpful contribution to the creation of a general catalogue though cannot be provided by this possibility since all exact title names would be required knowledge in advance.

A complete catalogue of games on the Twitch platform is thus difficult to determine. But how can we capture the total population of games as accurately as possible despite these complex restrictions?

Fortunately, in the present case, we can exploit certain features which are manifested in the market economy relationship between Twitch and Steam.

Looking at the customer journey of streamers, it becomes more than evident that a game has to be bought and subsequently played before it appears on Twitch as a stream. A game that has not been bought or is not played cannot be streamed on Twitch. Thus Twitch can be considered as an aftermarket product, or an aftermarket service provider for Steam.

Therefore only games that are sold or played on Steam can be available on Twitch. Of course, other video games may be offered for streaming, although such games are not relevant for the examined relationship between Steam and Twitch. We, therefore, consider the comprehensive product catalogue provided by Steam to be a sufficient population for subsequent analysis.

It, therefore, seems logical to start the selection and reduction of games on the Steam side before matching the thinner dataset to the games on Twitch.

We accordingly decided to enrich the games available on Steam with metadata from the open Steam library Steam-spy in order to qualify them for this project in two dimensions. To be included in the sample, a game should have had more than zero players on average in the last two weeks before data recording and prove more than 20.000 estimated sales. The level of 20,000 estimated sales represents the lowest possible resolution provided by Steam-spy, not allowing for a more precise limitation.

Altogether a total of 816 Steam games could be extracted on 03.03.2020 which should be examined in the following data recording.

Those 816 games were then merged and coupled with the corresponding games on Twitch. Specifically, this means that each game is assigned a unique ID by Steam and Twitch, but this ID is different on both platforms. The ID on one platform had to be matched and associated with the ID on the other side to determine the corresponding player and viewer numbers. In most cases, this assignment was achieved automatically via the name of the game, since Twitch, as mentioned above, provides an endpoint of its API for this purpose. This endpoint accepts a name and if an exact match is found, returns the Twitch ID. For most games, this process was feasible without errors with an exact name match for 600 titles.

Particular problems were encountered with Japanese and Russian titles which were translated into Latin characters. In addition, many games had suffixes or extensions to the actual game names. For example, Steam extends its product names with version information such as "Game of the Year Edition" or "Gold Version" which are not applicable to Twitch.

As a result, the games affected by these differences could only be assigned through traditional manual research using services such as Google or the dedicated Twitch search. As a result, a further 107 sample units could be obtained via conventional mapping. To avoid possible matching errors during this task, a dual review principle was applied.

Consequently, a total sample size of 707 matches was selected for which metrics should be recorded over the period from 03.03.2020 to 03.04.2020.

After these measures, the obtained sample size is compatible with the restrictions regarding the individual application layer interfaces and especially their rate limits as well as with the storage capacity of our query server and its backups locations. We expected the resulting data set to be approximately 10 Gigabytes in size.

Potential bias of the sample

It is to be questioned whether the selection of the cases based on Steam and the lack of specific consideration of the Twitch categories results in bias within the final data sets, that should be addressed at this point.

The selection criteria such as whether a game was played during the last two weeks as well as the minimum number of 20.000 owners might favour successful game concepts on the Steam Platform. While games with sustainable long-term design patterns are recorded, other Games that are played very rarely but still gather a significant number of viewers in Twitch streams are excluded from the data set. This requires particular consideration because such games show a specifically high viewer-to-player ratio that, purely arithmetically, tends to infinity. This characteristic suggests that, as outliers, they should be excluded as data points for further analysis one way or another during the data analysis.

We, therefore, assume that this extremely asymmetrical constellation in performance is of such unlikelihood that it renders no effect on the evaluation and the generalization of the data set. Games that are not played on Steam but reach high viewer numbers on Twitch are, also according to the post ante analysis of our data set, an extreme edge-case whose recording was negligible.

Data retrieval methodology

There are two fundamentally different approaches to storing information in big data projects. The first approach is to store the collected API responses unchanged in a straightforward fashion with later parsing and related organization of the data. The second approach parses the queried data immediately upon retrieval and thus makes it immediately verifiable.

The first approach is particularly suitable for projects showing alternating data formats and low output reliability. The direct saving of the responses without specific inspection of the returned data is characterized by a high error tolerance regarding the data collection. Any possible errors will only be dealt with during a data cleanup process which reduces the resulting inaccuracies. Complex checks of the data and sophisticated database storage processes are entirely avoided at this stage. This simplicity is both an advantage and disadvantage to this method. Because of the direct, simple storage, a logical check of the data concerning plausibility is not conducted, resulting in possible difficulties not being immediately noticeable during data collection.

In the present case of the data collection of Steam and Twitch, we already observed potential problems with this approach during the extensive planning process.

On the one hand, the existing data could only be accessed via interfaces that were undergoing major restructuring processes. Twitch is currently converting its old system over to the so-called "new Twitch API" while the old system is successively phased out. It was, therefore, advisable to use the new interface immediately to counteract potential problems, especially considering the long-term data collection. The reliability of this new interface, however, was at times only partially positive. Short dropouts and longer response times during our testing were not uncommon. The individual and specific obstacles will be described in more detail in section "Technical challenges using the Twitch API" below.

Also Steam had noticeable difficulties to answer all queries fast enough during the testing phase, especially during peak periods regarding player numbers. However, the dropouts were of such short duration that requests sent just moments later were already being answered again on both platforms. We, therefore, decided to implement a comprehensive error handling protocol to be able to collect all data even under heavy load on the services.

However, this decision does not come without a cost to the complexity of the query system. The implemented logic no longer allows queries against the service to be stored unchecked. Instead, it is necessary to parse each response directly, analyze it for the returned errors and respond to them appropriately. Thus, with regard to the possibilities mentioned at the beginning of this section concerning big data collection methods, we are focusing on the more complex of the two possibilities, which enables greater flexibility for error correction immediately within the query method.

Error prevention and correction measures

While the technology stack will be described in the next section, we would like to point out the implemented error correction measures without going into the technical background in too much detail.

First of all, we analyze every response sent by the servers for validity, plausibility and especially thrown errors. If the system detects a violation of these rules, it repeats the request up to three times. If the query still fails, the current game is skipped in the query logic, and the failed attempt is documented in a log file. We differentiate between "Steam exceptions" and "Twitch exceptions", meaning errors that occurred on one or the other platform. In this way, it is possible to react quickly in case of failure and isolate the specific problem. In addition, we gather metrics about the reliability of the services and therefore the reliability of the data collected.

Altogether, these intensive error correction efforts minimized the failure rate of purely data collection to 0.2%. In absolute numbers, this means that only seven out of over 3000 recorded points in time indicate technical errors and must be consistently excluded from the data record.

The second measure is more of an organizational nature to make the discussed errors visible to ourselves as maintainers of the process within a shorter time frame. If emerging errors cannot be handled automatically, they have to be analyzed manually and, if possible, corrected or at least evaluated. For this purpose, we implemented two separate and redundant systems. On the one hand, we were notified about the data collection processes and the cumulative number of errors via e-mail on a quarter-hourly basis. Over time, this allowed us to develop a solid understanding of how the scripts operated. We also received reports via e-mail regarding the daily running processes of database backups and import of Twitch Stream Tags. The former helped us to reduce the storage load of backup processes in several cases by compressing the exported files in order to accomplish everyday backups even towards the later stage of the data collection. In addition to this notification via e-mail, we also developed an online dashboard that provided us with all the relevant information regarding the current status of the data collection. In particular, we implemented diagrams to visualize the collected data in order to obtain a fundamental summary, including an opportunity for an initial inspection (figure 08).

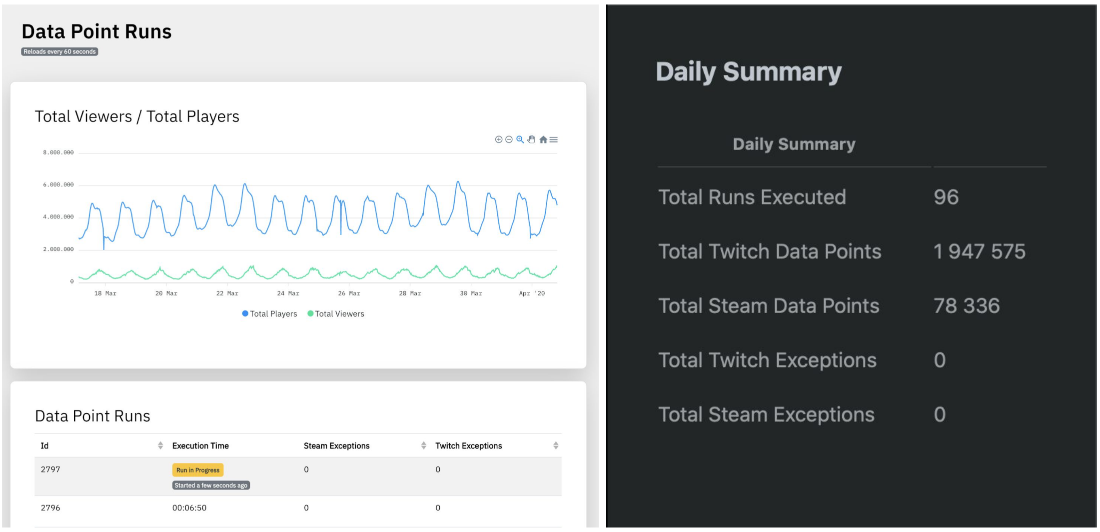

Figure 08. Reporting instruments during data collection

Technology Stack

The data collection process within data science projects often uses tools that are also employed to perform the actual analysis itself. Popular programming languages in this context are, for example, R or Python. Due to the widespread use of these programming languages in the data science community and the great experience of the users in this area, there are many extensions and packages for this type of data collection (See Munzert et al., 2014). Especially web scraping and text-based, but also quantity driven data mining approaches are easy to implement with this platform. Taking a first peek into the environment of data science, one can easily get the impression that there are not many alternatives to the widely discussed approaches.

Unfortunately, however, our experience with Python or R and the ecosystem that exists around it was rather limited. We implemented previous projects primarily in the programming languages Swift, Javascript or PHP. Consequently, from our perspective, it seemed natural to avoid learning a completely new programming language to build on existing expertise. After an initial orientation phase and evaluation of the available instruments, we decided to implement the PHP framework Laravel for our data collection software (See Otwell, 2016). The alternative frameworks and environments like Symfony (PHP), Django with Requests (Python) or other swift-related concepts were either too technically demanding or not well documented to be fully functional in the given time frame.

Laravel is lightweight enough to be adapted quickly and efficiently to the requirements of this particular case. Especially the object-relational mapping and the automation for creating models and controllers were decisive arguments. Many repetitive tasks that are required when using conventional frameworks are not required anymore with Laravel. In addition to this simplicity of operation, the applied framework is also sufficiently robust to reliably collect data over the set period of time. The package, which has now been released in version 7, can be considered a standard in the development of web applications. For the purpose of this research, however, we deploy version 6.17.1 in order to avoid potential bugs associated with the latest and thus not extensively tested version. The improvements and features included in the latest version would have held little to no relevance to the primary purpose of this research anyway.

Even regarding the use of extending modules, we tried to refrain from experimental versions and ideas by using well documented and maintained systems. We continuously executed this principle by using the HTTP client "Guzzle" (*Guzzle, PHP HTTP client — Guzzle Documentation*, n.d.), the Twitch API Client (Zipp, 2020) or the SQL Helper "Eloquent Join" (Horvat, n.d.).

For the database, we prefer to use the presumably most conventional option named MySQL. The fundamentally strong support and excellent documentation of this database allowed us not to hesitate in this decision. Lightweight alternatives such as SQLite would have been conceivable, but in connection with Laravel, they pose a challenge in terms of documentation and maintainability.

In retrospect, especially given the relatively minor number of technical errors during data collection, the choice of the technical foundation can be considered justified.

The completed data collection repository with all its features and dependencies is publicly available at GitLab.com for consideration and auditing (gitlab.com/cbs-master/cbs-master). The source code of the interface is provided in the same way and can additionally be accessed as an operational version at dashboard.cbs.lennartzellmer.de (gitlab.com/cbs-master/cbs-master-ui). The data collection code is furthermore part of the appendix of this thesis.

Technical challenges using the Twitch API

Obtaining the data from Twitch, in particular, was challenging us on three levels we want to address firmly to show which measures we applied and how they impact data validity.

Firstly the depreciation of the old interface with a subsequent transition to the new Twitch API, as already discussed briefly, posed minor complications towards our process to obtain a valid dataset. For better distinguishability Twitch names its API versions with nicknames which we will also refer to in this thesis in order to improve readability. The old version V5 which will not be supported in the future is called *Kraken* while the new and in the near future solely available interface is called *Helix*.

Twitch is not designing its own application programming interface specifically for scientific data collection, but rather for the various stakeholders and platform supplementors. These can cover the entire spectrum of tools and applications such as analytics platforms, plugins for computer games or supporting software for streamers. The challenge for Twitch, thus lies in building an API that meets the needs of all these stakeholders while offering the highest degree of flexibility in terms of scalability and interoperability. The boundary resources should thus enable the broadest possible innovation spectrum on the supplementing partners' side while protecting Twitch's core business (Bianco et al., 2014; Karhu et al., 2018). This is, at the same time, also the motivator for Twitch to constantly improve and further refine its boundary resources. While using Helix, it thus becomes evident that the overall usability has significantly increased compared to the old Kraken API. These improvements focus in particular on the method of authentication against the service as well as interacting with relatively high volumes of data. Special attention was paid to the request model of particular streams and their meta-information.

This development can also be seen through an increase in the rate limit, a factor that proved to be particularly advantageous for this research. While Kraken enforced a strict limit of one request per second, which would have been considered too slow for this data collection, Helix is now using a new points-based system.

For this, the service assigns a certain volume of points to each client, which can be exchanged for requests in communicating with the Helix API. Twitch initially grants each client 800 points of credit and charges them one point per request. New points are automatically added back to the account over time. If the balance drops below the maximum of 800 points, an additional four points are added to the client's account every second. If the points are used up, no more requests can be made until the account reaches a positive balance again (*Guide*, 2020). This results in an effective rate limit of 240 requests per minute for continuous use with additional 800 requests at the beginning of a request sequence. With a little over than 700 games in our sample, the rate limits of Helix could be considered sufficient while Kraken being slightly underpowered for our purposes.

After this decision for the Helix API, however, more challenges arose. While Kraken offered the feature to query the current accumulated number of viewers for each game, the new interface no longer provides such a possibility. Thus the information needed for this research was not immediately retrievable via Helix. That leads us to the second set of difficulties regarding the Twitch API: The complex retrieval process for the total number of viewers.

The only possibility to obtain the total number of viewers via the new API was to query all streams that are currently live for a game in their entirety in order to then add up the individual viewer counts. This seemingly straightforward task was, however, compromised by an incorrect implementation of the data pagination. Now, what is pagination itself?

If one searched on Twitch for example for streams about the game Fortnite, a very popular multiplayer game, one would receive about one thousand unique streams. Although one would receive a response from the server, the query would take a considerable amount of time due to the large amount of data searched for. Therefore well-structured APIs convert this data into chunks. Thus also at Twitch, it is possible to query these thousands of streams solely in blocks of 100. Each of these blocks again contains a cursor referring to the next block of one hundred items. This allows, for example, to query the mentioned 1000 games in 10 blocks of 100 items and thereby increase the speed of the application for the average user. Unfortunately, Twitch's resources suffer from a serious bug in the implementation of this design pattern. The error can be well illustrated by the 1000 games example. Instead of returning a correct empty cursor after the 10th request and thus indicating the end of the list, Twitch sometimes sends an invalid cursor that points back to the beginning of the list. Instead of creating a finite list, the result is an infinitely extended list that repetitively delivers the same content. (See *Cursor is not "null" anymore when there is no next page*, 2019). This false cursor did not follow any recognizable pattern, except that it was significantly shorter than all previous ones. The intermittent nature of the bug was not an advantage but caused major delays in the testing phase. A probe that checked the length of this cursor and in case of doubt aborted the query finally solved the problem.

The total amount of streams with all associated meta-information could be aggregated to total viewer numbers in the subsequent data analysis process.

Lastly, as the third complication, the exact mapping of the games on Steam and Twitch will be discussed in detail. After having queried the game catalogue of Steam and reduced it to a reasonable dimension one is confronted with the task to rediscover these games on Twitch in order to link the corresponding information.

Unfortunately, computer games do not provide globally uniform numbers or identifiers that allow a transparent verification of 100% consistency of these games. Some examples of such a standardized system would be the digital object identifier (DOI) standard in the scientific community or the International Standard Book Number (ISBN) used by book retailers. In these systems, a single number is sufficient to identify and retrieve a product across several distributors.

The game industry, and especially the platforms that rely on protectionism, have not yet been able to agree on such a uniform standard. What comes closest to such a common industry standard appears to be the independent database compiled by the community-driven gaming provider "Giant Bomb". Twitch even utilizes this independent database and provides the identification numbers they use for the ranked games. Unfortunately, Steam does not actively participate in the efforts of unification to increase interoperability and does not list this de facto standard with its products.

The only way to assign the games across the platforms thereby remains in the name of the software with optionally included subtitles. As already described, we, therefore, matched the games automatically if the names matched exactly. For the about 100 manually assigned games, visual matching criteria like cover artworks and screenshots were also considered. If there were large optical and game-logical matches between the titles in question, these were mapped accordingly. One of many examples of such a game is the title referred to on Steam by "古剑奇谭 (GuJian)". This game appears to be listed on Twitch in the Romanic alphabet under its reduced name "GuJian "showing visual similarities in the game itself as well as in the logo and title. The remaining games without a corresponding counterpart could predominantly not be identified at all or in some cases not conclusively. Due to a rather conservative evaluation, we only allowed an assignment in very explicit cases. Even with only slight signs of a doubtful match, no mapping was allowed.

Technical challenges using the Steam API

The Steam API was particularly demanding in terms of its documentation, formatting and the accessibility of information. It was not necessarily the lack of information that was problematic, but rather the fact that this information was poorly organized or difficult to retrieve.

With respect to documentation, the largest game-distribution platform on the market is not exactly the best example of detail and support. While other platforms maintain their boundary resources and see them as an opportunity, Steam shows little commitment in this area. The documentation websites are "manually maintained, rarely updated, and often don't reflect the current nature of the Steam Web API" (Skiles, 2020). This also reflects our experience in using the resources provided by Steam. Starting with incorrect descriptions of endpoints, unreliable game data, and even occasional total failures, the Steam API did not particularly meet our expectations.

Especially the formatting of information such as system requirements or release dates often did not follow a consistent standard. It can be assumed that developers provide these data independently as free text fields to Steam and are published without validation or further post-processing. Consequently, it was difficult to obtain machine-readable and analyzable data directly from the primary platform. The processing of these unstructured data sets into an analyzable format would have gone far beyond the time frame of this research.

As a technical aid, we, thus, collected wherever possible primary data from Steam itself and, in difficult scenarios, enriched this data with information from a meta-database called Steam-spy. This database aggregates and collects the data provided by Steam in various formats and provides it again in condensed and corrected machine-readable formats. As with any reduction and aggregation of data points, a certain amount of information is lost in this translation. We therefore randomly analyzed the data from Steam-spy by comparing it with the primary data from Steam and, except for the expected reduction in information, could not detect any deviation regarding validity. We, therefore, consider the data from Steam and Steam-spy as being interchangeable and use them equivalently in this work. By merging these two data sources, we were able to capture a comprehensive picture of the overall situation on the platform.

An overview of which data was obtained from which interface can be obtained from the table provided at the end of this section.

Technical challenges using the Steam-spy API

The only fundamental challenge in using the Steam-spy API involved frequent nightly downtime at the beginning of data collection. However, these failures resolved after a few days, so that the reliability of the service was not a fundamental problem in the final data collection. Beyond this small limitation, the meta-database proved to be very reliable.

An introduced rate limit of 4 requests per minute proved to be straightforward with our set goal of over 700 games. Our query logic even handles occasional errors with ease and allows us to keep the script executable even under heavy network loads.

Technical challenges collecting enriching data

In the following, enriching data especially refers to data as *system requirements* as well as *difficulty* and *multiplayer ability* of games. This data differs from the data collection discussed so far to the extent that it was not collected over a longer period of time. Instead, it was retrieved only once in its entirety from different service providers.

While the collection of the variables *difficulty* and *multiplayer ability* of the games raised no notable difficulties, however, it appeared that *system requirements* for games are neither centrally stored nor retrievable via Steam or Twitch. Steam principally provides system requirements that are specified by its developers, but these are neither standardized nor automatically interpretable. While some developers only specify a processor clock speed, others specify specific models and model versions. Other service providers also list such system requirements on a text basis and leave it up to the end-user to interpret them and to evaluate their PC system accordingly.

Only one of the providers we analyzed interprets the provided system configurations against a reference system and calculates indexes on the basis of the system requirements. This provider, gamesystemrequirements.com, maintains a website with detailed specification charts but does not provide a query-interface for automatic interactions. Hence, it was necessary to develop a so-called website scraper which extracts the data from the website and integrates it into the existing database.

The associated query logic was comparatively lightweight and could be implemented via the already applied PHP framework Laravel (See Weidner, 2020).

After successful extraction of the data, we unfortunately once again were faced with the challenge of assigning games between platforms without a unique identifier. The only matching characteristic was repeatedly the title, through which we were able to assign the so-called system requirement index to 511 exact name matches. If even one letter between the names did not match, if a comma was missing or if a hyphen turned out to be another separator, a relationship could not be confirmed. Developing a fuzzy matching mechanism to deal with these edge cases would be out of the scope of this paper.

Since there was also no additional information available besides the name on gamesystemrequirements.com, a further assignment of games was unfortunately not possible. Without the comparison of further parameters like screenshots, pictures or subtitles, it was not possible to carry out this task with the necessary confidence level.

Out of these 511 assigned titles, however, 42 did not provide a valid score although the underlying data was available. Either, only an undefined entry was available for these values, or the minimum requirements exceeded the recommended requirements rendering them unlikely. Since this data came directly from gamesystemrequirements.com, we cannot explain why these games lack a valid index. Therefore we have to exclude them from our sample and settle for a total sample size of 469 games.

Despite all these limitations and constraints while matching the games, we believe that the available sample size is sufficiently large to perform meaningful analysis.

Challenges caused by the coronavirus epidemic

Recent developments concerning the corona-virus and the COVID-19 epidemic also had an impact on our data collection process. On the one hand, the increased number of players and streams on the observed platforms during this time presented great challenges for our scraping and data collection script. On the other hand, the connectivity to the servers of Steam and Twitch was partially limited and led to exceptions and unreachable queries.

Especially during the evening of March 19th, the first Thursday of the enforced ban on going out in Italy, Spain, various eastern European countries, parts of Germany, France and extensive other restrictions regarding social life throughout continental Europe, an increasing number of data collection failures occurred. Moreover, due to additional curfews in large areas of the United States, Canada and South America, continued increases in the overall utilization level of the observed services developed noticeably.

While a total breakdown of the data collection script could be prevented from the outset by applying sophisticated query methods, one in every four to five data points was lost and could not be recorded throughout the time period of 2.5 hours between 10 PM and 0:30 AM.

During this time, the execution duration of the script unexpectedly increased above the previous maximum of 7 minutes to up to 13 minutes. This compares critically close to the allowed maximum of 15 minutes per data collection run. If a data collection run were to take longer than 15 minutes, it would overlap with the subsequent round and thus potentially collect data from one game twice at the same time. As a consequence, we optimized timeouts and concurrency of our source code to increase stability and to speed up the execution time. Especially on the side of the Twitch API, significant reductions of up to 50% could be accomplished while increasing reliability and scalability. The average execution time after implementation of these measures can be estimated at approximately five minutes. Peak times of up to eight minutes can occur, but were reliably processed and, in the case of failed attempts, repeatedly dispatched and reported.

The reasons for this increased complexity and added difficulty in data collection may be found in the increased use of the accessed services. In principle, the query logic behaves with a constant query time with the same number of games, but with a linearly growing duration with an increasing number of streams per title. Therefore, if the number of streams on the platform Twitch increases and this happens with a large number of games simultaneously, the overall query period will increase significantly. The runtime function of the query of all Twitch streams can be described in the common Big O notation as $f \in O(\text{games} \times \text{total Twitch streams})$. While the number of games was kept constant over the data collection period, there were significant variations in the number of streams. Between approximately 200,000 and one million streams could be observed simultaneously. Especially in the period after March 19th, this number frequently rose to the aforementioned maximum and decelerated the data collection to as slow as 13 minutes.

However, the increased load which users exposed directly to Steam and Twitch that day alone does not sufficiently explain the increase in query times to the extent documented.

An additional explanation may be found within the infrastructure Internet itself. As reported by the operator of the world's largest internet exchange point, DE-CIX in Frankfurt, Germany, gaming traffic increased up to 25% at the times in question. (DE-CIX, 2020) This probably led to pairing difficulties between the Nuremberg data centre used for this research and the servers of the providers, which were partially not immediately reachable. However, this behaviour normalized to a manageable level over the course of a few days and even eventually returned to comfortable levels.

In summary, the restrictions imposed by many governments on citizens have led to a significant increase in the consumption of observed services. This change in user behaviour was a major challenge for all participants involved in internet networking operations and eventually even led to the throttling of broadcasting services such as Netflix and YouTube as well as other social media platforms.

The validity of the collected data, nevertheless, could be assured by the modification of the source code on March 19th and therefore stays uncompromised on a technical level.

Data Analysis Methodology

In this section, we will describe the data analysis process, which includes a description of the theories describing the quantitative methods we use in this paper. We will structure the theoretical section of this methodological section from the hypotheses that we identified earlier in this paper so that the analysis of each hypothesis is described with a theoretical background for the methodology used. We will start with a description of the software used for analysis in this paper, which is Power BI and R.

Power BI and R

In this section, we will present our methodology concerning the technical aspects of our data analysis, which was conducted using Power BI and R.

Power BI is a program developed by Microsoft, as part of Office 365. It is a program specifically for data analysis, in which the user can transform, visualize, and analyze their input-data. Power BI accepts a wide range of different sources of input, ranging from websites to Excel-spreadsheets to SQL-servers. Setting up a proper Power BI report usually consists of three phases: 1) Transformation 2) Data-model 3) Visualization.

The transformation step happens inside Power BI's "Query Editor" where users use the M-language to modify, transform and shape their data so that it can be analyzed properly. The query editor includes but is not limited to query steps such as: Setting the data-type of each column, adding custom columns, renaming columns, replacing values, replacing errors, and mathematical operations. The query editor is set up so that the user can overcome data challenges such as handwritten data, text and numbers mixed in columns, and data stored in rows instead of columns, which is very hard to analyze within Power BI. In the data collected for this paper, we had no such significant challenges, as everything was hardcoded with data-type, in the correct columns and appropriately split up in the correct tables from the data collection stage. Within data analysis and visualization, there are two types of tables: Fact (FACT) and dimension (DIM). Fact-tables are any tables that contain numerical values regarding the exact data problem. Usually, these tables contain the dependent variables. In our data collection, the fact tables are the two tables containing the current player information and current viewer information about the games on the Steam and Twitch platform. The dimension-tables provide peripheral information about the fact-tables, which in our case are all the other tables which provide facts such as genres, publishers, and languages of the games.

The second step is to set up the data model within Power BI, which essentially is telling Power BI how to connect the different tables provided in our dataset, and which value to use as its unique identifier to look up data from other tables correctly. Our data model is depicted on this screenshot taken from our Power BI file (Appendix 2).

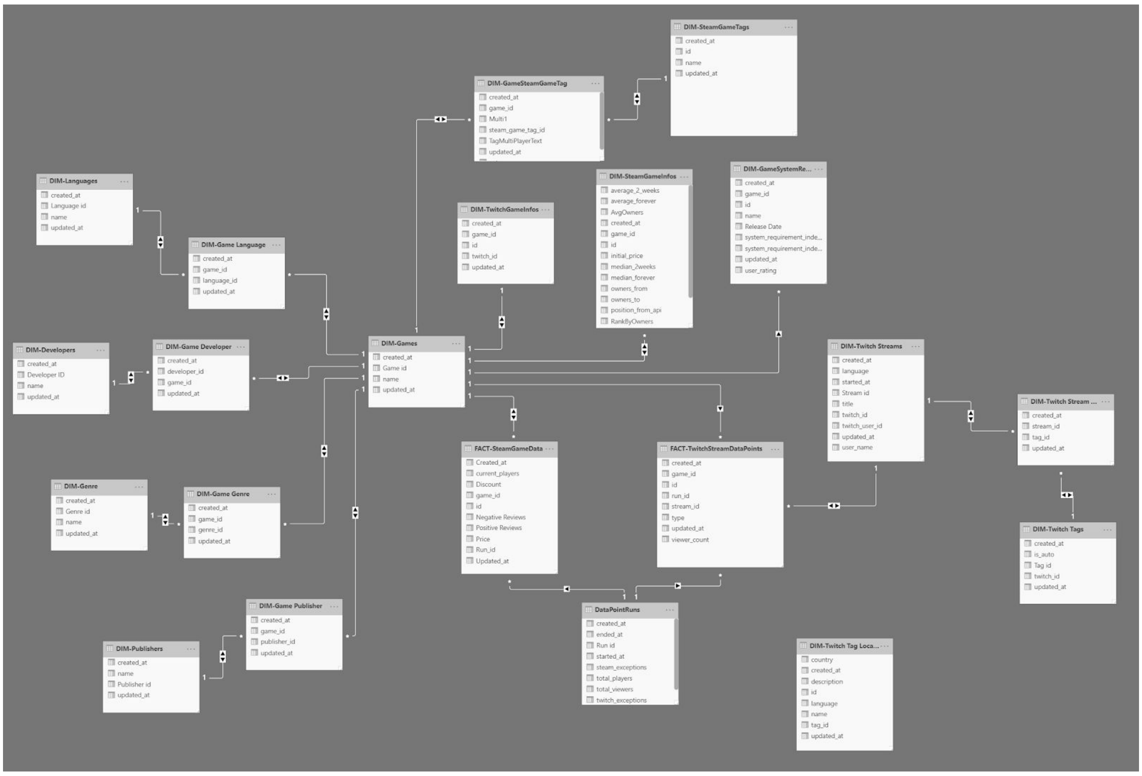

Figure 09. Data model in power BI

The data model made for this paper has Twitch on the right side of the data model, and Steam on the left. The two fact tables are connected via the “RunID” variable, which is listed in the DataRunPoint table. The Steam fact-table connects to the DIM-games table, connected by the unique “GameID” which is the identifier for the games. The DIM-games table is then connected further via “GameID” to the corresponding DIM-tables providing information about the games’ publishers, languages, game developers, and genres. Each of these DIM-tables has a corresponding lookup-table providing us with the text identifier for each corresponding unique ID in the first DIM-table. The DIM-games table is aditionally connected to the TwitchGameInfo and

SteamGameInfo which contains additional information about the games, connected by the unique "GameID".

The Twitch side of the data model is connected via further DIM-tables providing periphery information regarding the streams collected during the data collection.

The last step of analyzing our data is the visualization stage, which includes things like making charts and plotting graphs, and writing measures in Power BI's programming language "DAX (Data Analysis Expressions). As the charts made and measures written are specific to the individual hypothesis, we will provide the code and explanation for steps taken in this part of the process, in the section for each hypothesis. To analyze each hypothesis using multiple linear regression, we had to use the R-software, which is well known to anyone who has worked with quantitative methods before. R is embedded within Power BI and allows us to analyze the data through the computational power and plots provided to us by the R software.

The reason for using Power BI with R instead of just R-studio alone is that we felt more comfortable modelling and transforming the data in Power BI, as we had never used R before. Being able to calculate the average price and other standard arithmetic calculations in DAX as an example, was much more comfortable for us than attempting to learn the syntax of R and modelling the data in there from scratch. Thus we, like in the data collection, selected the tools with which we had previous professional experience rather than learn an entirely new language for this paper. We feel this approach secured a more reliable data collection and analysis.

Choice of statistical analysis

For this paper, we chose to use multiple linear regression, also known as ordinary least squares, to analyze our data and test our hypotheses. In this section, we will explain why we made that choice and elaborate on the alternative approaches that we could have taken instead.

To properly determine what type of statistical test is appropriate, we first summarize the contents and context of the hypotheses, we want to test statistically. We have multiple dependent variables, one for each platform, being players and viewers. Both variables are quantitative and continuous. We also have multiple independent variables, all of which are continuous and quantitative as well. We have one control variable, Corona, which is an ordinal binary variable. The hypotheses of this paper have the purpose of examining the relationship between a dependent variable and multiple independent variables. To simplify this, we want to understand whether the relationship is positive or negative between the independent variables and the dependent variables, and the intensity of that relationship compared to the relationship for the same independent variable on the other dependent variable. For example, we wish to understand whether or not a progressing product life cycle decreases players, decreases viewers, and whether or not it decreases viewers more than players. Thus we need the direction of the relationship and the compared intensity of the effect on the dependent variables.

With many continuous independent variables, a control variable and a dependent variable, there are mainly two ways of testing our hypotheses, Pearson's partial correlation and multiple regression analysis. Pearson's partial correlation tests the strength of a linear relationship between two variables after controlling for the effects of one or more other continuous variables. Pearson's partial correlation does not, however, make a distinction between dependent and independent variables, which would be a problem in our research design. We are not interested in the effect that the number of users for a game has on price, but only in the effect that price has on users (Field et al., 2012).

Multiple linear regression is used to predict the outcome of a continuous quantitative variable based on multiple independent variables. Multiple linear regression is an extension of simple linear regression which is used when only one independent variable is the subject of analysis. Besides being able to provide us with the direction and strength of the relationship, multiple linear regression also allows us to determine how appropriate our model is at explaining the variance in the data, measured by the r^2 (Field et al., 2012).

In testing our hypotheses, we feel that multiple linear regression fits better to our current statistical needs, and thus decided to use multiple linear regression as the statistical methodology for answering our research question.

Clarification on variables used in testing of hypotheses

Life cycle

H1: A progressing product life cycle decreases viewers more than players

H1a: A progressing product life cycle decreases players

H1b: A progressing product life cycle decreases viewers

We hypothesize that a progressing product lifecycle will decrease the players and viewers, but we expect it to reduce viewers more than it reduces players.

In calculating age of the game, we calculate the difference between the release date, and the end of the data collection stage (april 3rd) with the following DAX measure, which returns the age of the game in years:

```
AgeofGame = DATEDIFF(MIN('DIM-SteamGameInfos'[release_date]),DATE(2020,4,3),YEAR)
```

Social interaction

H2: Product social interaction increases viewers more than players

H2a: Product social interaction increases players

H2b: Product social interaction increases viewers

For this hypothesis, we used the variable “multiplayer” to test our hypothesis. The data we used to identify this variable came from the Steam-database and is a tag that can be voted on by players for each game. Thus, the players can vote on which tags define what the game is, which means that the more votes each tag has, the more likely the game is to be defined as a game type adhering to the tag. Multiplayer is one of those tags, and games marked with the multiplayer tags will indicate that they are indeed games where multiple players can participate and interact.

Multiplayer games are by far the most popular to both play and view in our data set, most likely because they center around primal human needs to socialize and continuously experience social interaction (Zagal et al., 2000). We argue in this paper; however, that multiplayer is not a binary variable but a continuous one. The reason is that the game’s rules and the game’s props and tools facilitate social interaction on different levels (Zagal et al., 2000). A game like DOTA 2 facilitates social interaction profoundly, as the game is close to unplayable as a single player. DOTA 2’s rules, props, and tools of the game directly facilitate social interaction through the game. Other games like ARK: Survival Evolved also have a multiplayer option, but still offer a lot of single-player content. The game’s rules and props and tools facilitate social interaction sometimes, but not always, and it is not required to experience the game. Thus, we argue that DOTA 2 and Ark: Survival Evolved are both multiplayer games, but that DOTA 2 is more multiplayer than Ark, which should be reflected in the data analysis. To best estimate this difference in multiplayer, we used the tags voted on by the players for each game and calculated how much multiplayer was voted for as a tag, compared to other tags for that same game. This calculation will give us an indication of how much multiplayer defines the game, compared to other tags such as action or adventure. The DAX-measure used was:

MultiPlayerContinuous =

***VAR** MPVotes = **CALCULATE**(**SUM**('DIM-GameSteamGameTag'[votes]),*

***FILTER**('DIM-GameSteamGameTag','DIM-GameSteamGameTag'[Steam_game_tag_id]="a4d7506c-78ca-473c-bff5-00a9b7108a2e"))*

***VAR** AverageTagVote = DIVIDE(SUM('DIM-GameSteamGameTag'[votes]),*

DISTINCTCOUNT('DIM-GameSteamGameTag'[Steam_game_tag_id]),0)

RETURN

***DIVIDE**(MPVotes,AverageTagVote).*

Access barriers

H3: Product access barriers increase viewers more than players

- **H3a:** Complexity decreases players
- **H3b:** Complexity increases viewers
- **H3c:** Economic risk decreases players
- **H3d:** Economic risk increases viewers
- **H3e:** Hardware requirements decrease players
- **H3f:** Hardware requirements increase viewers

H3a & H3b: Complexity decreases players & increases viewers

Complexity as a concept plays a significant role in game design. Difficulty as a sub-concept has many different meanings and can be used to incentivize players to behave a certain way and to increase the feeling of reward and punishment within the game (Juul, 2009). In essence, when a player plays a game, they always want to win. The win releases the endorphins and the dopamine, that rewards them for a task completed. However, if there was no difficulty and players are rewarded for simplistic tasks requiring minimal effort, the corresponding reward will be very low or non-existent. Difficulty serves the purpose of challenging players appropriately so that when the task or game is won, the reward is fulfilling enough. Too much difficulty can also be a problem though, as users feeling like the task is impossible or will require too much effort, will not be worth the reward in the end. Thus game developers have to balance on the edge of making the game difficult enough to be exciting and challenging, but also not too challenging as it can drive players away. The type of failure or difficulty is also important, as players prefer games where they feel responsible for the failure, as not feeling responsible for failure is associated with a negative perception of the game (Juul, 2009).

Difficulty is one of the determining factors as to how big the dopamine reward is for completing a task in a game. The corresponding in-game reward for this same task is also equally valuable to its degree of ability to be achieved. Wang and Sun (2011) argue that suitability of the reward to the level of difficulty is of utmost importance in game design, in order to attract both casual and more dedicated gamers to the game.

As with our hypothesis surrounding the multiplayer variable, we will analyze our hypothesis around difficulty as an entry barrier being a continuous variable calculated with the votes for the tag "difficult" compared to the average vote for other tags for that same game. We assume that the votes for this tag compared to the votes of other tags for the same game will give us a pretty accurate representation of a very subjective matter of difficulty. What is difficult for some may be easy for others and vice versa. The following DAX-measure was used to calculate difficulty as a continuous variable:

```
DifficultyContinuous =
VAR DifVotes = CALCULATE(SUM('DIM-GameSteamGameTag'[votes]),
FILTER('DIM-GameSteamGameTag','DIM-GameSteamGameTag'[Steam_game_tag_id]="330c1a0c-333c-4cbc-8cc6-3e5a11dc0dd9"))

VAR AverageTagVote = DIVIDE(SUM('DIM-GameSteamGameTag'[votes]),
                   DISTINCTCOUNT('DIM-GameSteamGameTag'[Steam_game_tag_id]))
RETURN
DIVIDE(DifVotes,AverageTagVote)
```

H3c & H3d: Economic risk decreases players & increases viewers

As previously explained, we use the price of the game to operationalize the concept of economic risk. With the price of the game, we stored the value in dollar cents. In interpreting the results, we felt it was easier to interpret the correlation coefficients as its impact on the game's players and viewers with the increase of the price by $1. Thus we chose to transform the data into dollars instead of dollar cents.

The initial assumption that the price of the game decreases players and increases viewers came from a previous paper we conducted. However, within this hypothesis, there could be concerns about reverse causation. Interpreting the results of this variable, there are in our minds two possible conclusions that could be true depending on the initial assumption. One could logically assume from a microeconomic theory perspective that price would cause player aversion and decrease players of the game. However, the relationship between price and players could also exist, that the games with the highest prices are the best and most popular games, and thus the price would have a positive correlation with players because it is affected by a third unquantifiable value along with the sorts of “entertainment value”. We assume in this paper the first assumption to be true, that price acts as an access barrier, and that an increase in price will be seen to decrease the players as fewer will want to spend the money to play the game. We also believe the price of the game to increase viewers, as players who cannot afford to play the game may alternatively choose to instead watch it be played on Twitch.

H3e & H3f: Hardware requirements decrease players & increase viewers

Hardware requirements are stored as an index in our data, from minimum to maximum. These numbers note how comparably heavy the hardware requirements of the game are on a scale of 0-100. To most fairly calculate the hardware requirements for each game, we assume from personal experience that playing the game on the minimum requirements is unpleasant, and there is no need for users to run the games on maximum output. We thus take an average of the minimum and maximum hardware requirements and accept that as the hardware requirements index for that particular game. We transform the variable using the following DAX code:

```
SystemRequirementsAvg =
((SUM('DIM-GameSystemRequirementsInfos'[system_requirement_index_min])+SUM('DIM-GameSystemRequire
mentsInfos'[system_requirement_index_max]))/2)+0.00000000001
```

We anticipate this variable like all other access barriers to increase viewers and decrease players. This anticipation is because we assume that an access barrier to playing the game, might be replaced with the alternative of watching the game instead. We believe the same to be true for hardware requirements as for the other variables in the access barrier hypothesis.

Requirements for multiple linear regression

Fields et al. (2012, p. 271) tell us that to conclude on our regression analysis, we have to check if certain assumptions are true about the sample. I will in this section go through those assumptions. I will also clarify which tests were performed to test these assumptions and thus, the validity of our model. Field et al. (2012) present the following nine assumptions regarding linear regression models, that should be met or corrected for, in order to present valid findings.

1. Variable types

The first assumption is that all variables used in the prediction of our dependable variable(s) must be quantitative or categorical. It should also be true that the outcome variable is continuous, quantitative and unbounded. What this means, is that both players and viewers must be measured in numbers, and have no constraints put on the data in the data collection, by for example only measuring games that have above a certain number of players.

Both average players and average viewers fulfil this assumption, as they are quantitative, continuous and unbounded.

The variables used for prediction are all measured continuously and quantitatively, except for the "Corona" dummy variable which is a binary variable measured as 1 for the data collected during the global lockdown, and 0 for the data gathered before the global lockdown.

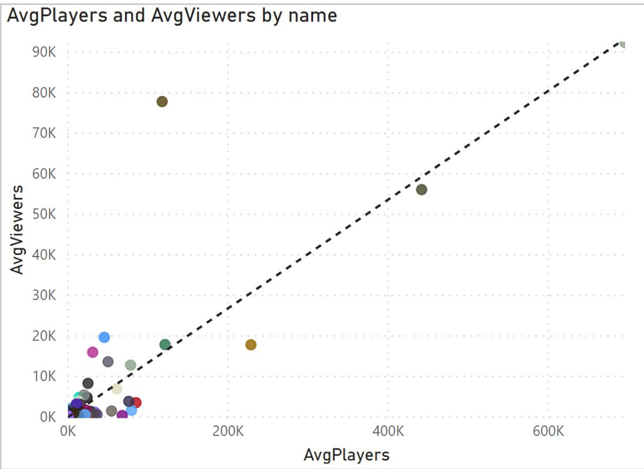

Figure 10. Average players and average viewers by name

2. Non-zero variance

The second assumption is that the predicting variables have variation in their value. All the predictors used in the model have variance in their outcome per game. The variance can be seen from the attached Power BI appendix in the tabs for each variable (Appendix 2).

3. No perfect multicollinearity

Testing the model for perfect multicollinearity is essential, as the predictor variables should not correlate too highly, as it would skew the accuracy of the model. A violation of this assumption would mean that our coefficients would be less likely to represent the entire population. A violation of this variable would also mean that the R-squared is slightly skewed. R^2 is a measure of how much of the variance in the dependent variable is explained by the independent variables. If our independent variables explain the same variance but are both added to the model, R^2 would be artificially increased. A violation of this assumption would cause concern in the analysis, as identifying the important variable of the ones that are correlated.

To test this assumption we follow Field et al.s (2012) methodology, and calculate the VIF values, 1/VIF and the mean VIF values.

The following code in R was executed for both players and viewers:

```
MLR <- lm(log1p(AvgPlayers) ~
MultiPlayerContinuous+AvgViewers+DifficultyContinuous+SystemRequirementsAvg+AgeofGa
me+AvgPrice, data=dataset)

library(lmtest)

library(gplots)

library(car)

sinkplot()

par(mfrow=c(2,1))

vif(MLR)

1/vif(MLR)

mean(vif(MLR))

sinkplot("plot")
```

The code created the following output, showing VIF as the top four values and the tolerance (1/VIF) as the bottom four values in the following output from R within Power BI:

VIF for players as dependent variable

```
MultiPlayerContinuous            AvgViewers  DifficultyContinuous
             1.179786              1.041405              1.102373
SystemRequirementsAvg             AgeofGame              AvgPrice
             1.660362              1.344277              1.611515
MultiPlayerContinuous            AvgViewers  DifficultyContinuous
            0.8476112             0.9602416             0.9071338
SystemRequirementsAvg             AgeofGame              AvgPrice
            0.6022784             0.7438943             0.6205340
[1] 1.323286
```

VIF for viewers as dependent variable

```
MultiPlayerContinuous            AvgPlayers  DifficultyContinuous
             1.181924              1.042239              1.102345
SystemRequirementsAvg             AgeofGame              AvgPrice
             1.662978              1.344899              1.610672
MultiPlayerContinuous            AvgPlayers  DifficultyContinuous
            0.8460777             0.9594726             0.9071572
SystemRequirementsAvg             AgeofGame              AvgPrice
            0.6013307             0.7435502             0.6208589
[1] 1.324176
```

Figure 11. VIF for players and viewers as dependent variables

Field et. al. (2012) says that there should be cause for concern if the largest VIF is greater than 10, if the average VIF is substantially greater than 1, or tolerance (1/VIF) less than 0,1 or 0,2.

Neither of these causes for concern hold true, and we therefore assume no multicollinearity in our regression model.

4. Predictors are uncorrelated with "external variables"

The fourth assumption is that no other variable should be able to predict the outcome of our predictors, as other variables not included in the model would be able to predict the outcome of our dependent variables.

While impossible to predict and test for every single variable that was not included in the model, one can assume no obvious variables left to test that would be able to predict the outcome of our dependent variables that were not included in the model.

5. Homoscedasticity

Homoscedasticity is the assumption that the variance of the predictor variables should be constant. If the residuals are unequal, there is heteroscedasticity. We test this assumption mathematically by performing the Breusch-Pagan test in R, which tests the null hypothesis that there is homoscedasticity in the model. We execute the following code in R, once for the players as the dependent variable and once for the viewers as the dependent variable:

```
MLR <- lm(log1p(AvgPlayers) ~
MultiPlayerContinuous+AvgViewers+DifficultyContinuous+SystemRequirementsAvg+AgeofGa
me+AvgPrice, data=dataset)
library(gplots)
sinkplot()
lmtest::bptest(MLR)
sinkplot("plot")
```

Breusch Pagan test for players as dependent variable

```
        studentized Breusch-Pagan test

data:  MLR
BP = 6.1325, df = 6, p-value = 0.4085
```

Breusch Pagan test for viewers as dependent variable

```
        studentized Breusch-Pagan test

data:  MLR
BP = 7.6388, df = 6, p-value = 0.2658
```

Figure 12. Breusch-Pagan test for Players and viewers as dependent variables

With a p-value of above 0,05 for both outputs we accept the null hypothesis that there is homoscedasticity in our dataset.

6. Independent errors

To test whether or not our residuals are independent, we will perform the Durbin-Watson test as described by Field et al. (2012).

The Durbin-Watson tests the serial correlation between the errors. It specifically tests the correlation between adjacent residuals (Field et al., 2012). The test results vary between 0 and 4, with a value of 2 corresponding to completely uncorrelated residuals. The test can return values between 0 and 4, with values below 2 corresponding to a positive correlation. According to Field et al. (2012) tells us that a conservative rule of thumb is to reject values less than 1 or more than 3, but be attentive around values that differ significantly from 2 depending on model and sample size. When performing the Durbin-Watson test in R, a p-value, which we want to be above 0,05 to show that our errors are independent.

To perform the Durbin-Watson test, the “lmtest” library and the required package “zoo” were installed in R. Following this, Field et al. (2012) instruct that we define our regression model and then run the Durbin-Watson test. The following code was run in R through Power BI twice, once for players and once for viewers as the dependent variables:

```
MLR <- lm(log1p(AvgPlayers+1) ~
MultiPlayerContinuous+log1p(AvgViewers+1)+DifficultyContinuous+SystemRequirementsAv
g+AgeofGame+AvgPrice, data=dataset)
library(lmtest)
library(gplots)
sinkplot()
dwtest(MLR)
sinkplot("plot")
```

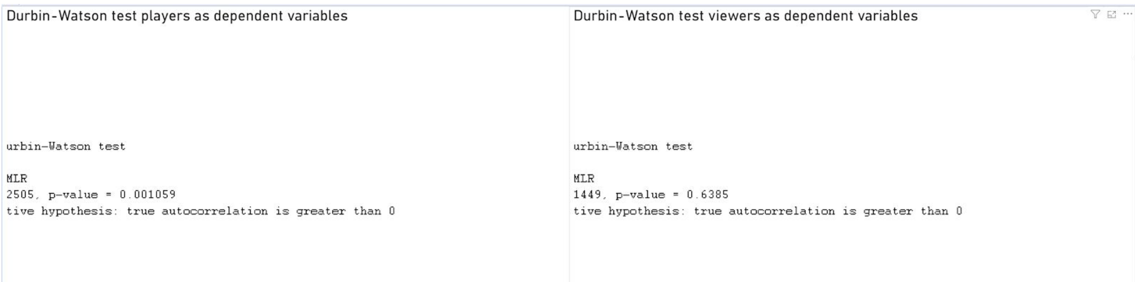

Figure 13. Durbin-Watson test for Players and viewers as dependent variables

The output that followed showed a Durbin-Watson value for players showed a Durbin-Watson value of less than 1, with a p-value smaller than 0.000, which means that with players as the dependent variable our errors are not independent of each other.

For viewers, the same code was run but replacing viewers with players and vice versa, and the output showed a Durbin-Watson value of around 1.86, with a p-value of 0.6385, meaning that the errors with viewers as the dependent variable are independent.

As a limitation to this study, we expected there to be autocorrelation in the data, as yesterday’s player numbers undoubtedly affect today’s player numbers, and we are not surprised to find that there exists autocorrelation within our data with the players as the dependent variable. We expected the same thing to happen for the data with viewers as the dependent variable. However, we believe that the reason that viewers have no auto-correlation is that Twitch is assumed to be highly event-based. Yesterday’s player numbers have no prediction power over whether or not a big tournament is hosted the next day, or if a popular streamer suddenly decides to stream one game over another. However, the difference in auto-correlation between the dependent variables is not an issue, as it can be expected in time series data for the players, and can be explained by the highly event-based environment that Twitch operates within (Field et al., 2012).

7. Normally distributed errors

The assumption that our errors are normally distributed is not to be confused with an assumption of the predictor variables being normally distributed. This assumption states that we assume the difference between the predicted variable and the outcome to be most frequently zero or close to zero. To assess whether or not our errors are normally distributed, we perform the Shapiro-Wilk test. In the Shapiro-Wilk test, the null hypothesis is that our errors come from a normally distributed population, and thus we are looking for a p-value above 0.05 to accept the null hypothesis (Razali et al., 2011). We ran the following code in R, once for players and once for viewers as the dependent variable.

```
MLR <- lm(log1p(AvgPlayers+1) ~
MultiPlayerContinuous+log1p(AvgViewers+1)+DifficultyContinuous+SystemRequirementsAv
g+AgeofGame+AvgPrice, data=dataset)
library(gplots)
sinkplot()
shapiro.test(residuals(MLR))
sinkplot("plot")
```

Shapiro-Wilk Players dependent

```
        Shapiro-Wilk normality test

data:  residuals(MLR)
W = 0.98016, p-value = 0.2422
```

Shapiro-Wilk Viewers dependent

```
        Shapiro-Wilk normality test

data:  residuals(MLR)
W = 0.98566, p-value = 0.5052
```

Figure 14. Shapiro-Wilk test for Players and viewers as dependent variables

The output showed a p-value above 0.05 for both players and viewers as the dependent variables, which means we accept the null hypothesis that our errors are normally distributed, and thus this assumption is fulfilled.

8. Independence

We assume independence in the model, which in our model means that each variable comes from a separate data run. We know this, as each data run has a unique identifier "DataRunID".

9. Linearity

We assume the relationship between the predictors and the outcome variables to be a linear one. In order to do linear regression, this assumption must be met, as it otherwise would skew the model and limit the applicability and generalizability of the findings of the model. We assess this assumption by looking at the "plot" output from R. We are looking for the red line in the residuals vs fitted plot to be close to horizontal around 0, to conclude linearity in the data.

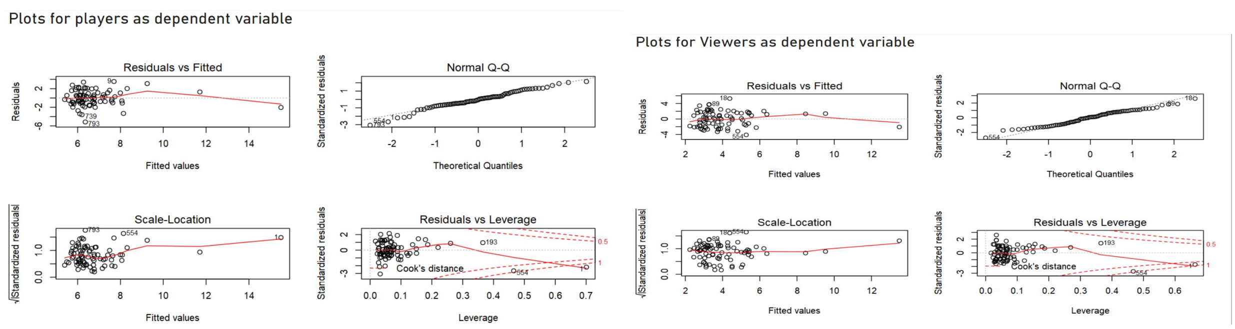

Figure 15. Plots for testing linearity for players and viewers as dependent variables

In both the plots we see linearity in the residuals vs fitted plot, and we thus accept that this assumption has been met.

Limitations

For this paper, we used a multiple linear regression OLS approach to testing our hypotheses and interpreted the results of the full model, including every variable. This way, we control for the variance explained by the variables in our other hypotheses, and can interpret the output in the context of all the factors we hypothesize can have an impact on the cross-platform behavior.

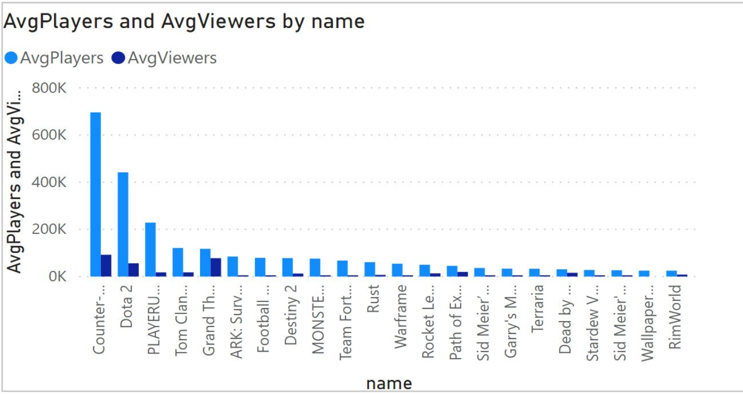

Figure 16. Bar chart of players and viewers by game name

Certain limitations naturally apply to the findings of this paper as a result of this methodology, which we will cover in this section of the paper.

OLS

OLS or ordinary least squares is the method we used to conduct our statistical tests in this paper. The methodology carries with it some limitations that we will explain in this section. The first problem is if there are major outliers in the data. The outliers can skew the prediction of the OLS model, and discard all the other values that are less significant, just so that the outlier value can be appropriately predicted (ClockBackward, 2009). As our data is not linear but follows a power curve as seen in figure 16, where a few games are much more popular than the rest of the games, the dependent variables follow an exponential function closer than a linear one. To properly analyze the impact through an Ordinary Least Squares (OLS) methodology, we do a log transformation of the data. Thus in our final models where we test the hypotheses, the code in r is *lm(log1p(dependent variable+1) ~ independent variable 1 + independent variable 2 etc*. This log transformation has some impacts on the findings and applications of this paper, which we will cover in the next section. OLS further has problems with accurate prediction when too many variables are added, or when the variables are not powerful enough predictors. In the final model used for results, we used all of our variables used in hypothesis prediction, because we felt like it was the most accurate tool for estimating the dependent variables. Our initial hypotheses were that all the variables had an impact, and in very few cases they came back with p values smaller than 0.05, which led us to include all the variables in the regression model. Our R2 was around 0.75 which is not abnormally inflated.

We acknowledge in this paper that these limitations apply; however, we argue that performing OLS regression is one of the more simple and basic statistical tools to use. Thus, with the extremely comprehensive data we have acquired, this is a first step in attempting to understand this data properly. Further research could then dive deeper into more complex and sophisticated statistical analysis on this dataset, but we believe that OLS is an appropriate methodology for the initial analysis of the phenomena we can observe through our acquired dataset.

We also believe that the non-linearity problem and issues with performing OLS on exponential data following a power curve was solved with our log transformations. We will cover the impact of log transformations in the next section.

Logarithmic transformation of the user interaction variables

Performing the log transformation of our data was a decision that was made after realizing that OLS cannot accurately predict data that is not normally distributed. Without the log transformation, many of the nine assumptions of linear regression would be violated. The violation of these assumptions led us to the conclusion that doing the model on the untransformed data was not viable for researching valid findings, and thus we decided to do a log transformation. The log transformation is the number to the 10th power that would equal the number it is given as input. For example, DOTA2 had 441.622 average players but had 6.65 log+1 players. In all of our models, we ran the model with this log transformation. The log transformation was log1p(dependent variable+1). We added a +1 within the log1p function, as our log was returning negative values for some of the games. The log function returned negative values for some games, due to an incredibly small number of average viewers and players for a few games. Thus we decided to add 1 extra average player and viewer to every game, to make it equal and not disturb the data too much.

The log transformation allows us to analyze our independent variables in an OLS model without violating essential assumptions for the testing. The problem is that the results of this OLS will not necessarily result in easily understandable coefficients, but rather in the effect of the variable on the logarithmic transformation of players and viewers. What this allows us to do, however, is see the nature of the relationship between the independent variables and the dependent variables (positive or negative) and the relative impact of the variables in relation to each other. For example, we will be able to test all of our hypotheses for this paper on the logarithmic data, as all of the hypotheses are exploring the cross-platform interaction, and not necessarily how *much* each predictor affects the result specifically.

We argue that for this particular paper the analysis done on the log-transformed data will still yield very viable results, and is truly the lesser of two evils when it comes to data violating the assumptions of the statistical tests. While it would have been preferable if we could have interpreted the coefficients in their purest form, we believe that transforming the data to log+1 allows us to answer our research question and test our hypotheses. This decision was clear for us concerning the research, but should still be noted for reasons of transparency.

Circular dependency between Twitch and Steam

Another critical methodological limitation is that there quite certainly exists a cyclical and circular relationship between Twitch viewers and Steam players. The general findings from our data suggest that the peak of Twitch viewers happens slightly after the peak of Steam players, suggesting a behavior of users to play some games and then watch a bit of Twitch streams afterwards. In order to appropriately analyze the cross-platform interaction between the two platforms, we need to analyze the impacts of our predictors on both players and viewers. While one might think that a ratio such as the player to viewer ratio might be the solution, this puts games with 400,000 players and 100,000 viewers at the same importance as games with four players and one viewer. This equalization is not representative of cross-platform interaction, and thus we cannot use this as an accurate measure for our research. Instead, what we have chosen to pursue is running the statistical tests twice, once with each platform's users as the dependent variables.

Because of the circular dependency and the finding that they most likely share the same user base, we have chosen to include each platform's users as the independent variable in the analysis of the other platform's users as the dependent variable. We made this decision also because one of the assumptions in linear regression is that there are no other variables that predict our result better than that of our current variables in the model. One of the best predictors of one of these platform's users is by the nature of their complementary relationship, the other platform's users. Thus for accuracy in the model and to control for the variance explained by the games' popularity on the other platform, we included them as independent variables in our analyses.

Corona

The current global pandemic caused by the COVID-19 virus, gave cause for concern for this paper, as governments around the world initiated national lockdowns to slow the spread of the virus.

While the virus has many issues on a human and societal level, it also has implications for this research. With many more people staying at home, there is very likely to be an influx of time spent playing video games, as even the World Health Organization has recommended it, as a healthy way to pass the time during self-isolation (Canales, 2020). We were quickly aware of this, and decided to control for the variance in our model caused by this pandemic, by making a dummy variable called “Corona”. The variable puts out a 1 if the data was collected after March 14th, and a 0 if it was collected before March 14th. We arrived at March 14th as the lockdown date, as it appears from this article by BBC News (2020), that March 14th is the most accurate date, on which most countries issued a nationwide lockdown. We had this variable as a control variable in every single model we tested, except for the ones where we tested our assumptions. The reason for this, was that the software had issues with the model, where it would be unable to compute the test results as it could not find the inverse of the value “0”, which was what Corona had as its output, around 50% of the time. Thus we decided not to include Corona in the tests, but in all of our models as a control variable, to account for the variance caused by this global pandemic, and the following nationwide lockdowns.

In interpreting the results of this test, we will look at the correlation coefficients as well as the p-values, which have to be less than 0.05 to be statistically significant (Field et al., 2012). If the p-value is higher than 0.05 we will declare the variable to be statistically insignificant in predicting the result and thus not have an impact on the dependent variable.

Results

In this section, we will report on the results of the multiple linear regression analysis we conducted following the methodology described in the previous section. The section will be structured by each hypothesis and its corresponding result including its statistical significance.

In order to answer the research question on factors influencing cross-platform interaction from users on complementing platforms we tested the following hypotheses:

H1: A progressing product life cycle decreases viewers more than players

H1a: A progressing product life cycle decreases players

Age of a game measured in years and average players were slightly positively correlated r(0.0628), p=0.000.

H1b: A progressing product life cycle decreases viewers

Age of a game measured in years was negatively correlated with average viewers r(-0.0562) p=0.000.

H2: Product social interaction increases viewers more than players

H2a: Product social interaction increases players

Multiplayer measured as a continuous variable and average players were positively correlated r(0.1831), p=0.000.

H2b: Product social interaction increases viewers

Multiplayer measured as a continuous variable and average viewers were positively correlated r(0.1490), p=0.0174.

H3: Product access barriers increase viewers more than players

H3a: Complexity decreases players

Difficulty measured as a continuous variable and average players was negatively associated r(-0.4920), p=0.000.

H3b: Complexity increases viewers

Difficulty measured as a continuous variable and average viewers showed a positive correlation r(0.7980), p=0.000.

H3c: Economic risk decreases players

Price of the game measured in USD and average players were slightly negatively correlated r(-0.0087), p=0.000.

H3d: Economic risk increases viewers

Price of the game measured in USD was positively correlated with average viewers r(0.0109), p=0.000.

H3e: Hardware requirements decrease players

Hardware requirements as an index showed no statistically significant correlation with average players, p=0.1566.

H3f: Hardware requirements increase viewers

Hardware requirements as an index showed a slight positive correlation with average viewers r(0.0114), p=0.000.

The models had 1522 degrees of freedom, as well as an r^2 in both models of around 0.75, which means our models explain around 75% of the variance in the users of both platforms.

With the tests using players as the dependent variable, Corona had a positive correlation with the players, r(0.1883) p=0.000. Average viewers had a strong positive relationship with players r(0.6512), p=0.000.

With the tests using viewers as the dependent variable, Corona had a negative correlation with the viewers, r(-0.1851), p=0.008. Average players had a strong positive relationship with viewers, r(1.109158), p=0.000.

Table 1: Viewers as dependent variable / Model Results

	Model 1	Model 2
(Intercept)	-1.647252 (0.000)***	-4.239330 (0.000)***
Corona	-0.136499 (0.000)***	-0185139 (0.008)**
Players per game	0.807904 (0.000)***	1.109158 (0.000)***
Multiplayer		0.149039 (0.017)*
Difficulty		0.798041 (0.000) ***
Price		0.010928 (0.000) ***
System Requirements		0.011462 (0.000) ***
Age of the Game		-0.056248 (0.000) ***
Number of Observations	816	816
R^2	0.5439	0.7518
Degrees of freedom	10179	1522

****p < 0.001 ** p < 0.01 *p < 0.05*

Table 2: Players as dependent variable / Model Results

	Model 1	Model 2
(Intercept)	3.96473 (0.000)***	4.136366 (0.000)***
Corona	0.18409 (0.000)***	0.188323 (0.000)***
Viewers per game	0.67327 (0.000)***	0.651215 (0.000)***
Multiplayer		0.183139 (0.000)***
Difficulty		-0.492082 (0.000)***
Price		-0.008755 (0.000)***
System Requirements		-0.002344 (0.156643)
Age of the Game		0.062899 (0.000)***
Number of Observations	816	816
R^2	0.5448	0.7559
Degrees of freedom	10179	1522

****p < 0.001 ** p < 0.01 *p < 0.05*

Discussion

In this section, we will provide a critical discussion of the results. We will also offer insights into the implications these results have for future academic research, as well as the managerial findings that can be implemented in practice. Furthermore, we will finally provide some limitations to the study and results in its generalizability, as well as suggestions for future work on this topic.

Results of statistical tests on hypotheses

In the results section above we reported on the outcomes of the statistical tests described in the methodology section, which constitutes the results of our paper. This section, however, does not describe what the implications are of these results, and how to properly interpret them in the context of the case companies, COVID-19, or any possible explanations for results that didn't confirm our hypothesis. Thus we will in this section go over the results of each hypothesis tested and explain and interpret the results in the context of our case study, as well as describe implications that these results could have further on the academic literature and digital platform field of research.

Lifecycle of the game

H1: A progressing product life cycle decreases viewers more than players

H1a: A progressing product life cycle decreases players

H1b: A progressing product life cycle decreases viewers

Considering our hypotheses, each of them consists of a necessary and sufficient condition to fulfil them. In the case of our progressing product life cycle hypothesis, the necessary condition is that both players and viewers correlate negatively with the age of the game. The sufficient condition to fulfil the hypothesis is that viewers correlate more negatively with the age of the game than players.

We reject the H1a hypothesis because the necessary condition is not fulfilled, with the correlation values 0.0628 for players ($p=0.000$) and -0.0562 for viewers ($p=0.000$). A progressing product life cycle does not influence players negatively but instead increases the total number over time.

Even though we were able to accept H1b as a necessary condition, we were not able to accept the overall sufficient hypothesis of H1. The necessary condition of H1a and b was that both dependent variables were affected negatively by our independent variable. Thus, as we were not able to fulfil both necessary conditions, we cannot accept the overall hypothesis of H1. A progressing product life cycle does not decrease viewers more than players, as players are not decreased.

We will now discuss why the positive correlation with average players should not be taken at face value. We believe that the results of this test and the positive correlation of the age of the game with average players, means that the age of the game is insufficient as an operationalization of the life cycle concept. The reason that we believe this, is that while the age of the game shows itself as negatively correlated with viewers, it does not with players, as we would have believed. We further believe that the insufficient operationalization stems from the fact that there in our data are two different classes of games:

1) The traditional game that can be completed, and thus has a life cycle.

2) A multiplayer service game, where the competition and multiplayer aspect is the key to the game having a long product life cycle.

The first class of games is a very traditional storyline, possibly multiplayer game, that has players play through a campaign or a story, and has players complete the objectives of the game. After said objectives have been completed there are often some peripheral objectives, but the game does not have the appeal to be played for years at a time by a single user. This is the game type that we thought of when developing the hypothesis, and thus expected a naturally deduced correlation coefficient to reflect the halftime-value of the game, as well as confirm findings of other literature on the subject, clearly identifying the age of the game to have an impact on the players of the game (Khomh et al., 2012).

The second class of games is what we classify as multiplayer service games, where the release date is often non-consequential to the game's popularity and is much more often than not defined by the network effects that characterize traditional digital platforms. We will later in this discussion go deeper into the topic of how this area of research can draw parallels to traditional sports, but for the sake of interpretation, it is easy to imagine this result in the context of traditional sports. The most popular traditional sports such as soccer (football), tennis, badminton, baseball, cricket, and many more, cannot be completed. The entire premise of the popularity of the sport is the continuous competitive scene, and the ability to play against equally skilled opponents under a strict rule-set that very rarely is changed from the original game's rules. This type of sports phenomena is also present in the games we see at the top of the player- and viewer charts. DOTA 2, CS:GO, and PUBG are very much popular because they offer this same framework to compete in, and thus the competition itself becomes the very appeal of the game. Because this type of games' appeal is so heavily defined by network effects and the previous and permanent existence of other players to compete with, it could potentially explain the deviations in the data on the players' side.

A possible solution to this problem in the data, showing us these two types of games, could be to control for the genre of the game. While we had data for all the genres of the individual games, we estimated the data to be insufficient at classifying games as the correct genre. This insufficiency is because the genre definitions in the video game industry are very broad and unspecific. Almost every game is an "action" game, and "shooter" could mean many different things within the context of the gaming world. Thus we decided to not control for the genre, as we could not possibly interpret the results correctly without any error. If the data from Steam had been better, we believe it would add value to further research to control for the genre of the game, as it could help explain the difference in the calibre of popularity between these two types of games.

In general, this result should be interpreted in relation to our research question, that the age of the game has a negative effect on the interaction on Twitch, but has a positive effect on the user base of Steam. We still argue, however, that the life cycle concept if properly operationalized, could affect cross-platform interaction negatively on both platforms.

We suggest that further research on the topic or dataset use split-sample analysis to properly analyze the effect of a products life cycle on the cross-platform interaction. We suggest more specifically, to run the models on the two classes of games identified in this section separately. One model should be run on the multiplayer service games like DOTA2, CS:GO, and PUBG and one model should be run on the traditional life cycle games such as Football Manager 2019, Football Manager 2020, & Russian Fishing 3 and 4.

Social interaction

H2: Product social interaction increases viewers more than players

> **H2a:** Product social interaction increases players
>
> **H2b:** Product social interaction increases viewers

In the case of our social interaction hypothesis, we again assess the necessary conditions that have to be fulfilled in order for us to accept our overall hypothesis. The necessary condition of H2a and H2b are that both players and viewers correlate positively with multiplayer. The sufficient condition to fulfil the hypothesis is that viewers are more affected by the multiplayer variable than players.

We are able to accept both hypotheses, as necessary conditions, with players having the correlation coefficient 0.1831 (p=0.000) with multiplayer, and viewers having the correlation coefficient 0.1490 (p=0.0174) with multiplayer. Both relationships are significant with p-values below 0.05, which shows a high level of statistical significance.

Even though we were able to accept both necessary conditions through H2a and H2b, we were not able to accept the overall sufficient hypothesis of H2. The correlation coefficient for players was higher than that for viewers, thus indicating that interaction on Steam is more affected by product social interaction than the interaction on Twitch.

In order to gain a better understanding of the situation, it is advisable to consider the necessary and sufficient conditions individually.

The results of the necessary hypothesis indicate a strong incentivization for cross-platform interaction and behavior for the multiplayer concept. We argue here that being multiplayer not only incentivizes cross-platform interaction but in some cases is the only way to truly take advantage of the features and possibilities of each platform. Something that was also mentioned in our previous work on cross-platform consumer behavior, was that merely having many players was sometimes not enough of an incentive for streamers to stream the game. Some games are better suited to streaming than others, and we argue that being multiplayer is a significant indication that the game is better suited for streaming than games that are less multiplayer or are singleplayer. To exemplify this, if the game is not very multiplayer friendly, or single player in general, the only interaction method the streamer has of interacting with their audience is through the Twitch chat. In a multiplayer game environment, the streamer has the possibility of also interacting with their viewers through the game itself. This increased possibility of interaction undoubtedly creates a stronger relationship between the viewers and the streamer, which again could be deduced to meaning a more popular stream.

Although our initial hypothesis predicted that viewers would be affected than players, the results stated otherwise. Overall, we believe that this shows that gamers might be more inclined to play a game that’s multiplayer, rather than watch someone play a game that’s multiplayer. The increase of utility on the player side, through the actual social interaction of the game, is greater than the utility gained through entertainment value added on the streaming side.

We previously argued that multiplayer was not a binary variable but rather a continuous one. In the same sense, we believe that some games offer more social interaction than others. Multiplayer can mean both playing *against* other players, and *with* other players, often simultaneously. Mostly we argue, the social interaction comes from playing *with* others and not *against* others, or more pragmatically from teammates rather than opponents. Something that could then affect the operationalization of social interaction through multiplayer are the games that are multiplayer, but in a one vs one format. These games do not offer the same level of social interaction as team-based games do, but are conceivably just as much a multiplayer game as the team games. One example of this is PUBG or other games in the battle royal genre, where players can choose to play alone or in a team. The interaction gained from playing in a team is much greater but is not reflected in the data. We believe further that this explains the results of the hypothesis. Utility gained from the social interaction when watching a stream is conceivably less potent than experiencing the social interaction yourself when interacting with one's own teammates.

With the classification of the games above, it also has some managerial findings for game developers attempting to break ground and reach the very top of the popularity of games. As previously mentioned all of the most popular games are multiplayer games and fall into the second class of games mentioned in the life cycle hypothesis. DOTA 2 and CS:GO have been around for quite a while, and are still very popular. The games that have been able to challenge them have all been big multiplayer titles, such as PUBG and Fortnite. To honestly expect to reach this plateau of games as a game developer, we believe that this data also shows that the next game to reach this level of popularity will be a multiplayer game falling into the second class of games as described previously.

We nevertheless believe further, that this finding of multiplayer promoting cross-platform interactions is an appropriate operationalization of its upper-level concept of social interaction. Social interaction, or interaction between users, is widely supported in the scientific literature on digital platforms, and in many cases, is precisely what the platform was built to facilitate and emphasize. This finding is thus confirmatory of the current status quo in the literature on digital platforms. Therefore we argue that this finding can be applied broader than our case study, to further the research consensus on social interaction and its role in consumer behavior on digital platforms.

Access barriers

For the access barriers hypothesis, we initially hypothesized that entry barriers would increase viewers more than players. Furthermore, we had six sub-hypotheses that consisted of three different entry barriers that we predicted would decrease players and increase viewers. In this section, we will first go over each of the three access barriers individually and then end with an overall assessment of access barriers as an overall hypothesis based on the results of the three sub-hypotheses. We will firstly assess the sub-hypotheses (H3a-H3f) as they jointly act as the necessary conditions for the overall hypothesis of access barriers (H3).

Complexity

H3a: Complexity decreases players

H3b: Complexity increases viewers

We are in the case of the complexity sub-hypotheses able to accept both hypotheses. Complexity was negatively associated with players with a correlation coefficient of -0.4920 ($p=0.000$) and positively associated with viewers, with a correlation coefficient of 0.7980 ($p=0.000$).

The results of the H3a analysis suggest that our initial assumptions about the motivations of players were indeed valid. We initially hypothesized that complexity would act as an access barrier to some users, as the game would be fundamentally too difficult for themselves to grasp or figure out. Because of this, we figured that some users might be inclined to go and watch someone more skilled than themselves on Twitch play that same game, to learn tactics and strategies about how to play the game. In the class of traditional games like mentioned in the life cycle hypothesis, this might be done through walk-throughs on Youtube, showing the player how to complete a certain level or how to figure out a particular puzzle. In the second class of games, which are by far the most popular, this usual approach with showing solutions might not work, as the players may need to learn fundamentally different things about the game. These learning objectives for the player could be both more skill-based, or more strategic. If the game was straightforward to grasp, further instruction on how to play or learning objectives might never be required or developed by the user at all. The more complex games, however, incentivize heavily to learn from others who are very good at the game, a need that Twitch as a platform fulfils very well. In the light of these facts, it makes logical sense that our results on this hypothesis correspond to the practical reason as to why we see difficulty have a positive impact on the number of Twitch viewers for a particular game.

We can further assume that our initial considerations regarding the H3b hypothesis are reflected in the data. We believe that the reason complexity reduces interaction on Steam, matches our initial hypothesized background. More complexity reduces player activity on Steam significantly and thus creates an access barrier in the sense of Hazée et al. (2017).

We believed that the various access barriers tested in this paper would decrease interaction on Steam yet increase it on Twitch. Because the platforms offer substitutable ways of experiencing the product we believe that if an access barrier is too big for the user to overcome, they will substitute their playing of the game with watching the game instead. Thus we would reasonably explain our results with the idea that if a game is too complicated for a user to play, they would instead watch others on Twitch play the game instead.

With these results, we can lastly speak to the access barrier's impact on the cross-platform interaction in relation to difficulty. It is clear from the results of the testing of this hypothesis that complexity will decrease the interaction on Steam, and will facilitate increased interaction on Twitch. These findings reveal that what is an access *barrier* on one platform, can be an access *facilitator* on the complementary platform if the platforms are coupled appropriately.

Economic risk

> **H3c:** Economic risk decreases players
>
> **H3d:** Economic risk increases viewers

In the case of the economic risk, we are able to accept both sub-hypotheses. Economic risk was negatively associated with players with a correlation coefficient of -0.0087 (p=0.000), and positively associated with viewers, with a correlation coefficient of 0.0109 (p=0.000).

Economic risk operationalized through price was the second concept that we believed to help define and operationalize the upper-level concept of access barriers.

In H3c we hypothesized through a traditional and well-accepted economic standpoint, that an increase in price would lead to fewer players for the game in general, and that games with lower prices would attract more players through the well-accepted relationship between price and demand. The results of the OLS analysis seem to support this theory.

There could, however, be a critical discussion conflicting with our findings and our initial hypothesis. When considering the microeconomic principles of increase in price, it is for the same product. An increase in price for games is an analysis across a very diverse category of products, with significant quality differences. Thus, the standpoint could be that the best and most popular games are also the most expensive, and that price is positively correlated with game quality and entertainment value. This variable, be it game quality or entertainment value, is incredibly challenging to operationalize, particularly with the dataset that we collected. One of the more reasonable ways of operationalizing this variable might in fact be by the player numbers themselves. This reasoning can be followed with a few essential assumptions. The first assumption is that user screen time is a finite resource, both for the companies, but also for the consumers themselves. Thus it is rationally assumed that the users enter into a utility optimization strategy, as to best spend their screen time between the available options. These options include watching TV, browsing the internet, or playing a video game. Between the video games available to the user, the user is assumed to play the game that gives them the highest amount of utility at that given moment. Given this logical and reasonable assumption, we could then deduce that the best games are the ones which provide the most utility to the users, and is thus measurable through its player numbers. Given these assumptions and deduced relationships between players and game quality, it could be reasonably assumed that there would be a positive correlation between price of the game, and the average number of players. In this context, the quite low correlation coefficient of -0.0087 can indicate that the pricing of games is incredibly efficient in this industry.

Another reasonable assumption regarding the efficiency of price setting could also be explained by the high degree of transparency that exists around the quality of the game. There could reasonably be a presumption of the idea of a self-fulfilling prophecy to affect the results. The phenomenon of the self-fulfilling prophecy is very well established in other products or industries (Wan & Yang, 2019). The idea is that people will assume that a higher price automatically means a better product, and thus the following popularity of the product is not a function of its actual quality, but rather mass psychology. We argue, however, that in the video game industry this phenomenon is not possible, because of the degree of transparency that exists around each game and its content. Before a game is even released, there are often alpha tests, beta tests and tons of reviews around the game. This transparency allows the user to make a much more informed decision of purchase in comparison to traditional products. This review process is also a reason that platforms such as Twitch are very popular, as it gives an unquestionable view of how a real person experiences the game, showing his or her experience of the game in real-time directly. One could logically assume from the positive correlation between price and viewers that this review process would be more vigorous for the more expensive games, warranting further research before purchase. This effect is also well documented in the decision-making literature (Lindblom, 1959). Charles Lindblom introduced in 1959 the “rational decision-making model” and the “muddling through”-model. The idea is that with full information about the product and its alternatives, the consumer should use the very detailed rational decision-making model to make the decision in question. However, this is not true for all decisions, as some decisions are too insignificant, and should be made using the “muddling-through” model, where the decision is based more on intuition and experience than any rational deductible analysis. The point to retrieve from this is that very early on in the literature, there was a distinction between decisions that required deep rational considerations before making the decision and decisions that did not. We argue that an increase in price increases the risk of making a bad purchase, and thus increases the need for careful consideration from the user. This all leads to further amplification of an increase in price to drive more traffic to Twitch as a platform.

In general, what can be said about this variable's impact on the cross-platform interaction between the two case companies, is that price is positively associated with average viewers, and negatively associated with average players. We believe that price shifts interaction from Steam to Twitch, which is how it as a variable affects cross-platform interaction. We believe further that our findings support the more micro-economical standpoint in the literature, that price acts as an access barrier, and that an increase in price leads to a decrease in players.

Keeping this in mind, we can move on to the interpretation of the result of the relationship of H3d between the price of the game and the game’s average viewers. There was a positive correlation between price of the game and the game’s average viewers, which was stronger than that between the price of the game and the players. The result confirms our initial hypothesis that the price of the game acts as an access barrier, and that consumers might increasingly choose to go to Twitch to experience the game if the price point is out of their reach. Assuming a positive correlation between price and quality, the most desirable games to the consumer may be out of reach price-wise, and thus further incentivize the consumer to experience the game on Twitch through someone else, rather than buying the game themselves. This shift in interaction between the platforms further adds to the idea discussed in the complexity hypothesis, of access barriers and access facilitators. The higher the price, the more interaction is shifted from Steam to Twitch, which would indicate to us that price acts as an access *barrier* on Steam, and an access *facilitator* on Twitch.

Hardware requirements

H3e: Hardware requirements decrease players

H3f: Hardware requirements increase viewers

In the case of hardware requirements we are only able to accept one of the two sub-hypotheses.

Hardware requirements showed no statistically significant correlation with average players ($p=0.1566$). Even though it was negatively correlated, we are not able to accept the hypothesis because our significance level is not high enough. As previously stated, we use a significance level of 95%, which means we would reject any statistical result showing a p-value of over 0.05. The significance level for this hypothesis is lower than 95%, and thus we are not able to statistically claim any significant relationship between the independent and the dependent variable.

Hardware requirements seem to be however positively associated with viewers, with a correlation coefficient of 0.0114 (p=0.000).

System requirements was the last operationalized variable that we combined with the other sub-hypotheses, considered to operationalize the access barrier concept. The idea when we initially developed this hypothesis is similar to the other access barriers, that very high hardware requirements might deter players from certain games and that this in return would decrease players and increase viewers. The increase in viewers would stem from users seeking to Twitch to experience the game by watching it, as they are not able to play it themselves. Some of the hardware requirements that could be significantly deterring are for example the requirement of a virtual reality headset, which only around 1% of Steam's customers currently have (Heaney & Hamilton, 2020). Other hardware requirements that could deter players could be a high level of the typical components that make up a powerful PC such as readily accessible memory (RAM), central processing unit (CPU) or graphics card. If the games require a very high level of these, we hypothesized that it would deter players from playing the game, and instead experience it through Twitch, driving traffic from one platform to the other.

The results of the statistical tests were as previously stated that hardware requirements on an index-scale had no statistically significant relationship with average players. The cause for this lack of association, however, may be explained relatively logically. We could reasonably expect system requirements to display some of the same tendencies of correlation between quality and game requirements, where the higher the requirements, the better the gameplay experience. As previously described in the introduction, game equipment is developing fast, and newer games have the ability to take advantage of some of the new technological opportunities presented to them, in the form of more powerful personal computers. We argue that one could make a reasonable assumption that games taking advantage of these hardware opportunities on average could deliver better game experiences, thus attracting more players and nullifying the access barrier effect that we expected the variable to have.

For hypothesis H3f, the results were unlike H3e statistically significant. This result means that hardware requirements had a statistically significant impact on the average viewers on Twitch. The results confirms our initial expectations that gamers choose to substitute the experience of playing the game with watching the game when faced with an access barrier they cannot surpass. Some gamers might be able to surpass this access barrier, but the utility gained from playing that particular game over other games is not worth the monetary investment in new equipment.

With these results in mind, we can now speak to the findings of this hypothesis in relation to our research question. The findings of the relationship between hardware requirements and the interaction on Twitch and Steam, tells us that hardware requirements is a variable that affects the viewers of Twitch, but not the players of Steam, and is thus not a variable that incentivizes cross-platform interaction.

In the managerial application of these findings, it is relevant to know how access barriers such as hardware requirements might affect the adoption of the game developers' game. If a game developer significantly increases hardware requirements, for example, there may be an irrational fear that the increase will lead to far fewer players than if they had lowered their hardware requirements. With these findings in mind, one could reasonably expect hardware requirements not to have a significant effect on average players and expect more viewers to come from this increase. There are naturally exceptions to this finding, as we would not argue that making a game exclusively for virtual reality, for example, would not affect its adoption rate. However, for the vast majority of games, we would argue that hardware requirements would not affect the adoption rate for players. Thus our findings would indicate that the video game industry seems to have struck an appropriate balance between player adoption rates and hardware requirements.

Overall assessment of access barriers

H3: Product access barriers increase viewers more than players

For the concluding evaluation of the access barriers, it is necessary to assess the individual six sub-hypotheses jointly and to summarise their results as a whole. The results of the analysis of the three operationalized variables of price, system requirements and difficulty must be evaluated jointly in order to draw conclusions about the relevance of access barriers.

In this context, it should be pointed out that nearly all six sub-hypotheses had significant effects on the respective games or players. The only exception to this rule applies to the insignificant hypothesis H3e, which assumed a connection between hardware requirements and players. All other hypotheses can be seen as confirmatory of our overall hypothesis H3 for access barriers as an upper-level concept.

Although we did not observe a negative trend of increasing hardware requirements on player numbers, the fundamentally presumed effect of access barriers regarding user-interaction with platforms seems to exist. However, the observed access barriers differ fundamentally from the factor of life cycle previously examined. While the life cycle has an equally positive effect on both interaction patterns towards Steam and Twitch access barriers behave in the opposite direction on the platforms. Increasing complexity, economic risk and hardware requirements decreases interaction with Steam and increases it on Twitch. This result leads us to believe that variables that fall within the category of access barriers will shift the interaction from one platform to the other, acting as an access *barrier* on one platform, and an access *facilitator* on the other.

It seems counterintuitive that a concept we researched as access barriers leads to an increase of interaction on the Twitch side acting as an access facilitator. We believe this stems from the fact that we analyzed product characteristics on complementary platforms but often draw on the findings to conclude on them interchangeably. We do this because we believe that there is no true interaction on either platform without the products. It is hard to imagine the interaction on Steam without the user playing a game, and likewise on Twitch, without the user watching a stream. Because the platforms are built up so heavily centered around the products, we believe that we can make statements about the findings of this paper on the platform side, even though we researched the product side from a user perspective. In the instance of access barriers, however, we need to differentiate between the platform perspective and the user/product perspective. We argue that a concept that is purely an access barrier to the product from the user perspective can act as an access facilitator for one of the complementary platforms, from the platform perspective. When talking about the individual user; however, we do not find it reasonable to argue that price is a facilitator for that user's product engagement on the platforms.

The term of access facilitator we believe should be defined from its counter-definition in access barriers. This is because, by the nature of the term as used contextually in this paper, it can only exist from the existence of an access barrier. The access facilitator is the other side of an access barrier, increasing user interaction on the platform that is not negatively affected on its user interaction by the access barrier. Thus, an access facilitator leads to increased user interaction on a complementary platform, if its complementor suffers decreased user interaction because of an access barrier to the users.

The findings for this concept are both applicable in managerial and academic contexts. For the managerial side, it is crucial knowledge when making decisions, that one knows the potential impact of those decisions. For a game developer, implementing access barriers such as complexity, price, and hardware requirements, are very operational decisions that can both affect the quality and adoption of the game. If a game developer makes the decision to reduce the hardware requirements but sacrifice game quality, they might make that decision with the idea in mind that the trade-off is worth it and will increase the adoption rate. Our data suggests that the core decision to be made in this instance should be aimed at striking an appropriate balance between limiting access barriers on Steam, and the marketing that the game gets from Twitch. Striking such a balance is the key because our data shows that access barriers do decrease the interaction the game has on Steam but increases it on Twitch. Developers are interested in interaction on both, but are reasonably more concerned with the interaction on Steam, as that is where their customers pay for the actual product.

The academic findings of the impact of this concept on cross-platform interaction is also fairly interesting. The main contribution to the area of platform research within platform ecosystems is the role of access barriers in complex platform ecosystems, and particularly in the relationship between two complementary platforms. If one perceived the ecosystem that Steam and Twitch interact in, as two dimensional as proposed by Hoelck and Ballon (2015), this relationship would be closest to a vertical or diagonal relationship. We were able to identify access barriers as a defining concept for the interaction between those platforms.

Further research could benefit from looking at other ways of operationalizing these concepts to see if the results are still consistent with the findings from this paper. If the results are not, it could indicate that there are different factors within access barriers that explain the cross-platform interaction better than access barriers. Further research could also explore if the relationship between complementary platforms we found with the access- barrier and facilitator concepts still hold true for other variables within the access barrier concept.

General findings

In this section of the discussion we will critically discuss and summarize the findings of this paper, as well as attempt to interpret and explain the more general findings that were found as a result of this research, as compared to the hypothesis-specific findings that were just discussed.

The first fundamental contribution this study is able to contribute can be seen in the indications that Twitch and Steam both actually share the same user base. We see this from the significant correlation between average players and average viewers, and the similarities in response to the same regression variables. This finding is not particularly surprising to anyone who knows the platforms, but we believe it is a significant finding nonetheless, as it proves this assumption to be accurate, and can help further research in the area with a starting point. Admittedly, we can only provide an indication of this correspondence and cannot clarify the degree of conformity between the two user groups. Nevertheless, we believe that we are contributing towards a future proof in a positive way.

COVID-19

In this section, we will cover the general findings that came out of this paper, from the COVID-19 pandemic. As previously described, a lot of the data sample (March - April) was collected after most governments announced a lockdown. This, of course, put more people at home looking for entertainment, and fewer people spending their time with friends or family, or outside for that matter. To control for this variable, we created a dummy variable that showed whether the data was collected after or before the governments implemented nationwide lockdowns in most countries. We initially realized this would have an effect, and thus decided to control for the virus via our dummy variable. This paper was written during the lockdown, and it is thus also just interesting to find out what kind of effect this variable had on the cross-platform interaction in general.

The most apparent finding concerning corona was that user interaction on both platforms increased during corona. Players and viewers were as expected artificially inflated due to people staying and working from home. This shows through the very small p-value of the dummy variable “Corona”, showing statistical significance for players. Corona was negatively associated with viewers, indicating that Twitch did not benefit at all compared to Steam. The explanation for this, we believe to be quite natural and reasonable. The individual time investment one has to put in to play one game of Counter-Strike Global Offensive, or DOTA 2 is quite big compared to the time investment of watching someone play a part of a game of those same games. Twitch is available anywhere, on the phone and while cooking something for ten minutes. In general, the access barriers to Twitch are smaller. During the lockdown we assume that fewer people are transporting themselves, fewer people are outside of the house in general, and more people have the time to sit down and play a full game of something from Steam. This means that in general people have substituted Twitch slightly with Steam. We believe this also shows that most of the time, if possible, people prefer to experience the game by playing it themselves and not by watching someone else. The explanation for a negative correlation between viewers and corona, still causes some concern, however. We know from the data that the average viewers did increase in the same timeframe. However, when performing a regression analysis on viewers as the dependent variable and corona as the independent, it shows as not statistically significant $p=0.571$. It also shows a positive correlation in this test, indicating that most of the variance and increase in Twitch viewership during the COVID-19 pandemic, comes from the increased number of players.

Another explanation for this phenomenon could also be that many of the newer players are not the “regular” gamers. These are players that do not normally sit down to play, but in light of the quarantine and government lockdown, have decided to pursue this form of entertainment. These people might not even know or have the need for Twitch, as they are not very regular players, who do not benefit from many of the aspects that Twitch offers as a platform, in relation to the games. This would result in an influx of players to Steam, but the same amount of viewers for Twitch, which could also explain the difference in statistical significance for the dummy variable. We cannot conclude for sure on which explanation holds true for the effect of this pandemic, but we believe it is along these lines based on the results of the statistical testing.

Another thing we found concerning COVID-19 is that the relationship between the platforms did not seem to change from a descriptive statistics point of view. As we can see from the graph below, there seems to be a very cyclical relationship between the platforms.

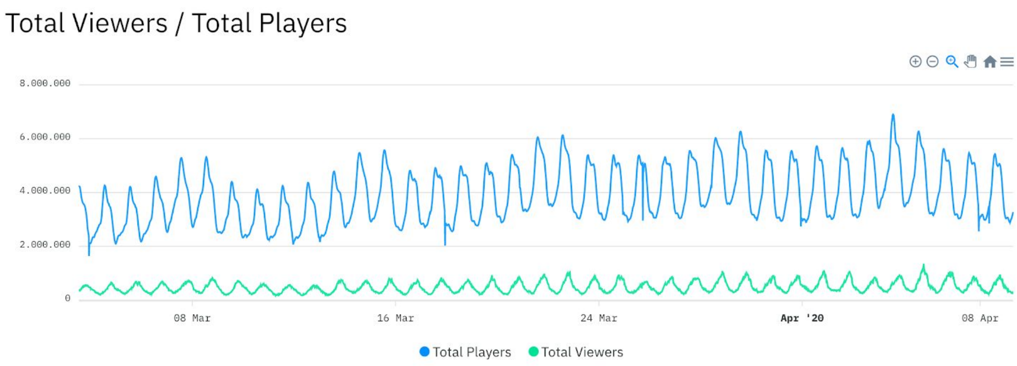

Figure 17. Total viewers and total player during data collection period

The cycle is one that is fairly natural to the daily operations of both platforms. There is a general peak of players and viewers, presumably when people around the world have the most free time and availability to play. The peak of Twitch happens slightly later than Steam, which indicates that people tend to play a bit, and then watch a bit after an ended playing session. This relationship has remained stable through the government lockdowns that happened as a result of the COVID-19 pandemic.

To summarize this section on the impact of the COVID-19 pandemic on this study, several things have happened to the data that we can conceivably measure.

- The number of players has significantly increased during the lockdown
- The number of viewers has increased, however mostly explained by the increase in players
- The cycle between Twitch and Steam has remained stable throughout the lockdown

We believe this area has significant potential for future research, using the same dataset or an extended one. Several more things could be researched concerning the difference between before and after lockdown, and this dataset presents a unique opportunity to research this difference, that might not come ever again. Questions that could be conceivably asked about the difference in consumer behavior could, for example, be aimed at the type of games people play. Maybe the lockdown has increased the screen time spent on more immersive games, or on games that generally take a longer time. In general, we believe that there are a lot of possible findings that could come out of this dataset, with a research question aimed at exploring consumer behavior on these two platforms, but the findings are unfortunately outside the scope of this paper.

Limitations

In this section, we will discuss the limitations that we find to be present for the findings of this paper. We will exclude the methodological limitations that we described in the methodology section. However, they still apply to the findings in general, but do not need to be further elaborated in this section. We will finish with a discussion about whether or not we believe the findings to be relevant despite these following limitations.

The first significant limitation we find to apply to the results of this paper is the fact that we are only analyzing two cases out of an industry. As described in the case selection, it was a reasonable and arguably logical choice to select these two cases, as they fit within the scope of the paper and were the respective market leaders in their category. This does not, however, free us from the limitation that we do not have data from other game providers such as Epic Games (Fortnite), Battle.net (World of Warcraft), Sony (Playstation), or Microsoft (Xbox). We also do not have the data from Twitch's two biggest competitors Mixer and Youtube Gaming. These other companies still provide significant data for cross-platform interaction but fell outside the scope of this paper. We believe that although more data could have been attempted to be gathered, the data collection for these two platforms was already time-intensive, and thus a decision was made to generalize our findings from the results of these two cases.

We believe that more generalizable results could have come from including all the companies within our industry, but we still believe that our findings are relevant and generalizable. Flyvbjerg (2006) wrote in his paper that one of the misunderstandings about case-studies is that you cannot generalize from a single case study. He argues that it is possible to generalize from a case study, granted that the rest of the parameters around the study are set up appropriately. We argue that we from our case selection and data collection stage, have selected appropriate cases, and collected appropriate amounts of data to be able to generalize our findings towards more general concepts such as cross-platform interaction in other digital complementary platforms.

Another conceivable limitation to the study is that it was conducted during the COVID-19 pandemic. We described most of the findings that came out of this as a result in the previous section, and we believe that the virus in general opened up for more findings and research than it limited the study. While the argument is valid that COVID-19 inflated players, which could have affected the reaction the variable had to different regression variables, we argue that after controlling for the pandemic through a dummy variable, it is no longer a limitation. In fact, it opened our eyes to a whole new area of research that can be conducted based on our dataset as previously described, and it could mean that future research would be able to investigate this cross-platform interaction amid a global pandemic. We argue that through the discoveries made in the context of the pandemic, and with a control variable for the player increase during lockdown, that COVID-19 should be seen as a feature of this study, and not a limitation.

In this context, it should also be noted that the data collection phase was only performed for one month. The time it took to set up the data collection stage, and perform the necessary analyses as well as discuss and conclude on the findings, made this limitation unavoidable. The video game industry is often seasonal depending on releases of big games and tournaments. Like most retail businesses, Christmas is the most active time for releasing products, and there is a certain expected break following Christmas because of this. With the nature of the timeframe in which we had to conduct this study, we have fallen into this break-phase, and it might have influenced the findings. However, we believe that the nature of the relationship between the platforms is so circular and interdependent, that we do not expect the variables to have affected the cross-platform interaction differently because of this limitation. Although the coefficients might be different during another time of year, we believe the nature of the relationship between the independent variables and the dependent variables remain the same throughout the year, nullifying this limitation in regards to our research question. Further research could benefit from conducting a full year of research, such that the general cycle of the industry is accounted for in the data collection.

This paper has another natural limitation related to the general scope of research. The hypotheses tested were developed from previous work done by the authors, and while adapted for this paper, we did not test any other hypotheses, other than those generated based on this previous work. This presents itself as a limitation to the study, as there are undoubtedly many more variables that could accurately predict or affect cross-platform interaction. Some of these variables we did not have access to, and could thus not test in our paper. Other variables such as game genre, or specific streamer characteristics, and the effect of the supply of streamers on demand of games, we had access to but did not test. These areas would most likely provide significantly different findings than those that came as a result of this paper, and thus would have yielded a different conclusion. We argue that these hypothetically different findings are not necessarily better than those found in this paper, just different. If the scope of this paper had been greater and more comprehensive, we would have included more general and larger hypotheses that required more in-depth statistical testing to answer. However, for the scope of this paper, we believe that our hypotheses and statistical testing were appropriate and that the findings of this paper remain relevant, although more findings would have added additional value. Further research could benefit from looking at these relationships, and possibly explore areas that we never considered for this paper, such as language and region data, or whether or not the individual game companies strategically design their games to be more enjoyable for spectators. If further research were to look at the relationship between some of these unexplored variables and their impact on cross-platform interaction we believe it would add to the academic findings of this paper, and thus not need to stand alone as its individual study.

Connected to the just discussed limitation of scope, we present the limitation that we used a purely quantitative methodology in the analysis of our research question on cross-platform interaction. While the data collection and the research question heavily favored a quantitative approach, we believe a mixed-method approach could have been viable as well and added additional credibility and explanation for the findings of this paper. A mixed-method approach with added interviews from a large enough sample of users would have explained more about the way that each factor affects the cross-platform interaction, and could have possibly helped with interpretation of the quantitative results we found. The scope of getting a reliable enough sample of users for both interviews, we believe would have been too extensive for this paper. As preferences and behavior vary heavily between users, the big data approach we took in this paper would have been excessively challenging to supplement with interviews, as the interviews would undoubtedly be biased towards availability of respondents. What we mean by that, is that the interviewees we would have been able to get in contact with, would inevitably not be representative of the entire population of users on Twitch and Steam, and thus not necessarily add that much value compared to the number of hours spent on acquiring the data. Having personal experience with the platforms ourselves over a long time, we believe that we were able to see the logical pragmatic flaws within the data ourselves and make adjustments, without having to rely on external input from interviewees. We thus argue here, that while a mixed-method approach would have been a viable option for answering the research question, the scope of this paper would not have allowed for the gathering of a representative sample to significantly add value to the quantitative results that we found through our big data approach.

Overall we argue that these limitations exist and affect the paper as a whole, but that most of them are limitations that occur as a result of the scope of the project. We believe that our findings in total are relevant and contribute to the current scientific literature on digital platforms in general, but mainly to cross-platform interaction- and gaming-related research.

Generalizability of our findings

In the following section, we will analyze to what extent our findings can be applied to other industries and scenarios inside the area of platforms. We will make comparisons between our case study and other industries that we believe are similar in enough ways for our findings to be viable enough for application in such scenarios. Hoelck & Ballon (2015) introduces the model for how a platform-dominated ecosystem looks, as previously introduced in the theory section of this paper.

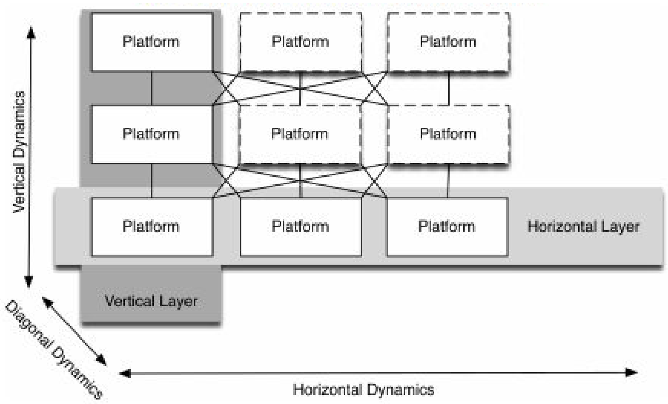

Figure 18. Model of a platform dominated ecosystem (Hoelck & Ballon, 2015)

The horizontal layer is about competition between platforms on the same market. This layer would be Mixer and Twitch or Steam and Epic Games Store. The vertical layer offers complementarities, while the horizontal layer offers substitutes. While we do not believe that our research contributes to this model directly, we believe that it can help us understand the dynamics and coupling between platforms in relation to our findings.

Other platforms offering the same services as Twitch and Steam are affected by the findings of this research. Twitch's competitors such as Mixer or YouTube Gaming, are comparable to Steam's counterparts like the Epic Game Store or Origin. In the following, it is rather necessary to analyze to what extent the observed coupling of the two platforms in the ecosystem of the video game industry is transferable to other industries and platform arrangements.

In the context of these digital platforms, we further examined the specific case between a distribution platform and a marketing-financed media platform throughout this study. The main challenge facing this case relates to the extent to which the specificity of the case applies to other areas of platform ecosystems.

A possible abstraction of this case could be conceivable through the joint product of the two platforms. We assume that the interaction between two platforms originates in particular from the products they jointly operate on and on which they offer different sets of services. Twitch and Steam are not the only two platforms that engage in such a relationship, and we believe our findings to be relevant for application in certain scenarios where a similar mainly digital relationship between platforms can be found. For this purpose, we will identify alternative purely digital platforms to which the research findings could be applied.

Among pure digital platforms, we believe these findings to be relevant to the other platforms on the direct horizontal and vertical layers to Steam and Twitch as portrayed by Hoelck & Ballon (2015).

One such example is the case of Netflix and IMDb that share the same product of movies. However, they are able to add value to the user experience through different complementary services. Netflix offers consumers the opportunity to stream movies and shows for a fixed subscription cost every month. IMDb offers users a one-stop information platform with a large variety and quantity of information around movies, shows, and actors while users will use both platforms, to add value to the same product. While the former platform distributes its product through its streaming capabilities and thus works analogously to Steam, IMDb generates reviews and summaries based on the movies. The monetization of IMDb is advertising-driven and can therefore also be seen in direct association to Twitch.

Similarities in the digital sphere are furthermore evident regarding the platforms DAZN and Onefootball. Founded in 2016, the sport live streaming service offers sports enthusiasts the opportunity to watch internationally popular sports without limits for a monthly subscription. It combines a variety of well-known sports leagues on one platform and offers a wide range of popular programs such as NFL, UEFA Europa League or NBA. Thus, it represents a platform that distributes the product of sport streaming in particular on the foundation of football (soccer). Onefootball, on the other hand, considers itself a media publishing company and, with their app of the same name, represents the very core of mobile European football media. On the media side, they likewise compile all major football leagues on a single platform and distribute their content through the use of advertising. Both platforms are thus based on the same product, the media broadcasting of football. It seems quite conceivable that similar effects in terms of social interaction and access barriers influence the users' connection to the platforms. The idea of the life cycle, on the other hand, appears to be hardly operationalizable in this industry, as the life cycle of live broadcasting events seems finite by the end of the broadcast.

Apart from these purely digital platforms, however, in the digital environment, also platforms based on real physical products can be coupled. Therefore and additionally, the question should also be answered as to whether coupled platforms with physical products can display similar effects for themselves or exploit them commercially.

Finding platforms that share the same product is relatively tricky outside of the digital world. The sharing of physical goods such as cars, tools or other common everyday products is in practice considerably more complex and demanding than in the digital domain. Therefore, the phenomenon is rarely seen in the platforms that operate outside the digital content and software sphere. Due to the digital nature of our original case, it is, unfortunately, unlikely to identify strictly analog cases in the physical product environment. However, by broadening the scope and looking at the wider spectrum for applicability of the results, it is possible to discover applications in the areas of influencer marketing or digital distribution platform of conventional goods.

Such an example of a digital distribution platform and its coupled complementary platform can be seen in the particular case of drop.com and YouTube. Drop is a crowdfunding network that differs from the market leader Kickstarter in the sense that it is specially tailored for so-called group buys. In these group buys, interested individuals unite to have a specific product manufactured on-demand, which would not be financially viable without the group-by for the individual participants due to high setup costs. The Crowdfunding Network thus manages a community of enthusiasts on the one hand and companies, designers and creators on the other. To reach a minimum order quantity, network and group order, participants rely on social networks such as Twitch and YouTube to talk about the new products available and acquire potential participants. In this context, they act like micro-influencers for their specific niche market. Since the products are mostly highly specialized custom-made solutions for the HIFI sector or the very specialized and small market of mechanical keyboards, the community members are, for the most part strongly emotionally connected with the products. This strong involvement with the products, similar to that of computer games, leads to a strong use of YouTube, Twitch and Discord for organizing and sharing knowledge. It is well conceivable that barriers to access to the products have a similar effect here as with purely digital assets. Even social interaction and the life cycle of the product can be applied in a similar way as for the purely digital domain. Although the example is very specific, it shows that similar behaviour can be observed in the case of physical products.

Distancing ourselves even further from the analogy of the platform's shared products, two things become clear. On the one hand, this allows us to see the limits of the applicability of the results, but also the boundaries of the adaptability with regard to the platform networks developed by

Hoelck & Ballon. We find that the complexity between platforms within platform ecosystems sometimes exceeds the two-dimensional perception.

An example of this overly simplistic view can be seen in a case where platforms share the same industry but offer different services and products. Flixbus and Uber is one such example. Both platforms offer transportation of consumers from A to B but in mainly different use cases and forms. Uber offers the service of individual transportation in small, partly privately-owned vehicles, thus differentiating itself from the long-distance bus service Flixbus. When looking at the platforms on a superficial level of transportation, one might thus be inclined to determine that they are platforms existing on the horizontal level of substitutes in the platform ecosystem. Both platforms offer a solution to the problem of transport between two cities and are therefore not complementary but substitutes and competitors in their market. We believe, however, as the platforms also offer fundamentally different products, they also exist partly on the vertical layer with each other. Flixbus offers long-distance transportation by bus of consumers from two predetermined destinations that users can buy tickets for. Uber also offers transportation of users, but usually for shorter distances. Uber offers the consumer to be picked up very close to their current location and being transported to their end destination completely. Thus with these two platforms in mind, one could easily imagine a consumer being transported a long distance with Flixbus and then purchasing last-mile transportation to their final destination through Uber. The coupling of the two platforms in this fashion, therefore, does not suggest a horizontal dynamic but rather a vertical to diagonal relationship depending on consumer usage. This multidimensional linking of platforms combining complementary properties with substitutes exceeds the explanatory power of platform ecosystem networks.

This complexity is also visible in the applicability of our results to the example of Uber and Flixbus. It is difficult to imagine that any access barriers to Flixbus would have an effect on the use of Uber in the complementary relationship. The same applies to the concepts of lifecycle or social interaction. The applicability of our findings, therefore, reaches its limits in this case.

This realization of the complexity within platform ecosystems also helps us to understand the findings of our paper in relation to our case companies. While we still believe the most accurate description of the relationship between the case companies is that they are complementary platforms, their dynamic is more complex than such. As Netflix stated in their 2019 letter to their shareholders, the main metric they are interested in is consumer screen time (Netflix, 2019). They believe that one of their main competitors is Fortnite, even though one would not believe Fortnite to be one of the biggest competitors to Netflix, because we often think of platform ecosystems like traditional markets as depicted in microeconomic literature. Twitch and Steam, like Netflix and Fortnite, also compete for consumer screen time and are thus both substitutes for entertainment, which is also evident in the findings related to our hypothesis on entry barriers. Trying to depict this complexity of being complementors and substitutes at the same time, while relying on each other is not possible in a two-dimensional model. We argue that the findings of this paper should all be thought of in light of this complexity of platform ecosystems, as we believe network effects and platform competition, substitution, and complementarity influences the individual findings heavily.

While the inherent nature of the complexity gives a good example of the boundaries of application, we were nevertheless able to identify several relatable areas of application. We furthermore discovered the boundaries of the findings by looking outside the idea of a shared product. However, this boundary is not clearly defined. Even less similar cases can be of great interest in this area. A just slightly similar case that might be less obvious to the research of platforms is the similarity between the popularity of the suppliers on Twitch and other media platforms such as Instagram. Media platforms, where the content suppliers act as influencers, function the same, and often pay their content creators in the same way. The bigger the supplier's platform is, and the more consistent user interaction they can generate as a content creator, the more they are paid. Our concepts that were statistically significant at predicting user interaction on Twitch may also be useful at predicting the user interaction on other media platforms.

Conclusion

The research area of platform ecosystems is slowly emerging from a niche perspective to an increasingly relevant research domain. In particular, the growing interconnection of platforms of industry leaders as well as multi-platform companies are contributing to this trend. Constant innovation in this field, together with the growing relevance of resource boundaries, further contributes to this development. Eight out of the ten most valued companies globally employ a multi-platform strategy, which means that they own and operate several digital platforms within their business. For these companies, it is vital to understand the optimal interaction of their ecosystems, consisting of highly complementary products. While the area of dual-sided markets and single platform mechanics has been very well researched, the research of mechanisms in platform ecosystems and complementary platforms is much less mature. We argue that our conclusion and findings from this paper contribute to a more developed understanding of these platform ecosystems and particularly to the mechanisms between complementing platforms.

Through a quantitative case study in the video game industry, we were able to answer the research question and ***identify factors that influence cross-platform interaction from users on complementing platforms.***

Firstly we found that a progressing product life cycle influences cross-platform interaction on two complementary platforms. While the results of this hypothesis were slightly different than expected, having the life cycle positively correlated with players, this led us to an important distinction and finding of this paper. We argue that there in the video game industry exists two different classes of games, the traditional life cycle game, and the multiplayer service game. We conclude that for other industries, the particular characteristics of the life cycle of the product should be taken into account when trying to research the user interaction on complementing platforms.

Secondly, we were able to identify social interaction as a significant factor in influencing cross-platform interaction from users on complementing platforms. Social interaction on the product increased the user interaction on both platforms in our case study, confirming the current status quo in the literature and also our expectations in the development of the hypothesis. We assume that the factor of social interaction shows a high degree of generalizability, primarily because of the overall network characteristics of platforms. The findings were confirmatory of our expectations and the literature, suggesting a more profound and deep connection to consumer behavior in general.

Furthermore, we were able to identify access barriers as a factor that influenced cross-platform interaction from users on complementing platforms. Furthermore, we examined the unique effects of access barriers in the relationship of complementing platforms. We found that access barriers act as an access barrier on one platform and an access facilitator on the other platform. We conclude that this behaviour is not a direct contradiction to complementarity, but rather evidence of the complexity that exists within coupling of platforms in complex platform ecosystems. We identified the sub-concepts of complexity, economic risk, and hardware requirements as jointly representing the overall concept of access barriers. We find the operationalization of the concept to be appropriate, but we acknowledge that the concept is not limited to these sub-concepts, especially in other industries and markets. We do expect, however, that the findings from this study on access barriers would see replication in research on other access barriers in the same context.

We can finally conclude that the relationship between two platforms in the complementary sense is not one-dimensional. Operating in a heterogeneous ecosystem, platforms often are not able to act purely complementary, leaving room in the academic literature for a model that more appropriately visualizes and explains this complexity.

In this paper, we were able to show that the complex nature of complementing platforms is a wide-ranging domain. This complexity originates mostly from the environment of platforms in general. The environment of platforms are continually changing, due to exponentially increasing technological capabilities, as well as social factors impacting platform networks such as COVID-19. This complexity and continually changing environment makes platform networks a fascinating topic for academic research, as well as an essential domain for industry leaders within platform markets.

Bibliography

Adner, R., & Kapoor, R. (2010). Value creation in innovation ecosystems: how the structure of technological interdependence affects firm performance in new technology generations. *Strategic Management Journal*, *31*(3), 306–333.

Allen, R. G. D. (1934). A Comparison Between Different Definitions of Complementary and Competitive Goods. *Econometrica: Journal of the Econometric Society*, *2*(2), 168–175.

Andronico, M. (2019, July 10). *How to Stream to Twitch*. Tom's Guide; Tom's Guide. https://www.tomsguide.com/us/how-to-stream-to-twitch,news-21077.html

Arthur, W. B. (1989). Competing Technologies, Increasing Returns, and Lock-In by Historical Events. *The Economic Journal*, *99*(394), 116–131.

Bailey, D. (2018, March 22). *With $4.3 billion in sales, 2017 was Steam's biggest year yet*. PCGamesN. https://www.pcgamesn.com/steam-revenue-2017

Baldwin, C. Y., Woodard, C. J., & Others. (2009). The architecture of platforms: A unified view. *Platforms, Markets and Innovation*, *32*. https://books.google.de/books?hl=de&lr=&id=1BvhQT8SHZkC&oi=fnd&pg=PA19&dq=The+Architecture+of+Platforms:+A+Unified+View&ots=uiBx9WCRHY&sig=fJavg30CptvOcd3FW0-RsNTSx3M

BBC News. (2020, April 7). The world in lockdown in maps and charts. *BBC*. https://www.bbc.com/news/world-52103747

Bianco, V. D., Myllärniemi, V., Komssi, M., & Raatikainen, M. (2014). The Role of Platform Boundary Resources in Software Ecosystems: A Case Study. *2014 IEEE/IFIP Conference on Software Architecture*, 11–20.

Caillaud, B., & Jullien, B. (2001). *Chicken & Egg: Competing Matchmakers on the Internet*. Mimeo, CERAS and IDEI, 2001 (b).

Canales, K. (2020, April 1). The WHO is recommending video games as an effective way to stop the spread of COVID-19, one year after adding “gaming disorder” to its list of addictive behaviors. *Business Insider*. https://www.businessinsider.com/who-video-games-coronavirus-pandemic-mental-health-disorder-2020-4

Carbaugh, R. J. (2016). *Contemporary Economics: An Applications Approach*. Routledge.

Chen, M. K., & Nalebuff, B. J. (2006). *One-Way Essential Complements*. https://doi.org/10.2139/ssrn.937384

ClockBackward. (2009, June 18). *Ordinary Least Squares Linear Regression: Flaws, Problems and Pitfalls | An analysis of the defects of least squares regression. | ClockBackward Essays*. Clockbackward.com. https://www.clockbackward.com/2009/06/18/ordinary-least-squares-linear-regression-flaws-problems-and-pitfalls/

“Complement.” (2020). Oxford University Press.

Consulting.com. (n.d.). *How To Make Money On Twitch: Everything You Need To Know*. Consulting.com. Retrieved March 24, 2020, from https://www.consulting.com/how-to-make-money-on-twitch

Cooper, C. L. (Ed.). (2015). Complementary Products. In *Wiley Encyclopedia of Management* (Vol. 46, pp. 1–2). John Wiley & Sons, Ltd.

Cournot, A. A. (1897). *Researches into the Mathematical Principles of the Theory of Wealth*. Macmillan.

Cumming, D. R. S., Furber, S. B., & Paul, D. J. (2014). Beyond Moore’s law. *Philosophical Transactions. Series A, Mathematical, Physical, and Engineering Sciences*, *372*(2012), 20130376.

Cursor is not “null” anymore when there is no next page. (2019, December 8). Twitch Developer

Forums.
https://discuss.dev.twitch.tv/t/cursor-is-not-null-anymore-when-there-is-no-next-page/23372

DE-CIX. (2020, March 19). Big upswing in Internet usage due to COVID-19 measures. *De-Cix*. https://www.de-cix.net/en/news-events/news/big-upswing-in-internet-usage-due-to-covid-19-measures

de Reuver, M., Sørensen, C., & Basole, R. C. (2018). The Digital Platform: A Research Agenda. *Journal of Information Technology Impact*, *33*(2), 124–135.

Edwards, C. (2013, November 4). Valve Lines Up Console Partners in Challenge to Microsoft, Sony. *Bloomberg News*. https://www.bloomberg.com/news/articles/2013-11-04/valve-lines-up-console-partners-in-challenge-to-microsoft-sony

Eisenmann, T., Parker, G., & Van Alstyne, M. (2011). Platform envelopment. *Strategic Management Journal*, *32*(12), 1270–1285.

Eisenmann, T., Parker, G., & Van Alstyne, M. W. (2006). Strategies for two-sided markets. *Harvard Business Review*, *84*(10), 92.

Evans, D. S., & Schmalensee, R. (2005). *Paying with Plastic: The Digital Revolution in Buying and Borrowing*. MIT Press.

Evans, D. S., & Schmalensee, R. (2005). *The industrial organization of markets with two-sided platforms*. https://www.nber.org/papers/w11603

Fahey, M. (2009, August 21). *Upcoming Blizzard Battle.Net Feature Draw From Warcraft, Xbox Live, Life*. Kotaku; Kotaku. https://kotaku.com/upcoming-blizzard-battle-net-feature-draw-from-warcraft-5342994

Farrell, J., & Katz, M. L. (2003). Innovation, Rent Extraction, and Integration in Systems Markets. *The Journal of Industrial Economics*, *48*(4), 413–432.

Field, A., Miles, J., & Field, Z. (2012). *Discovering statistics using R*. Sage publications.

Flyvbjerg, B. (2006). Five Misunderstandings About Case-Study Research. *Qualitative Inquiry: QI*, *12*(2), 219–245.

Fortney, L. (2018, October 5). *How Amazon's Twitch Platform Makes Money*. Investopedia; Investopedia. https://www.investopedia.com/investing/how-does-twitch-amazons-video-game-streaming-platform-make-money/

Frank, A. (2018, December 4). *Epic Games is launching its own store, and taking a smaller cut than Steam*. Polygon; Polygon. https://www.polygon.com/2018/12/4/18125498/epic-games-store-details-revenue-split-launch-date

Game System Requirements. (2020). *Game System Requirements*. Game System Requirements. https://gamesystemrequirements.com/faq

gamingandtechnology. (2008, March 13). *Steve Jobs introduces the App store - iPhone SDK Keynote*. Youtube. https://www.youtube.com/watch?v=xo9cKe_Fch8

Gawer, A. (2014). Bridging differing perspectives on technological platforms: Toward an integrative framework. *Research Policy*, *43*(7), 1239–1249.

Gawer, A., & Cusumano, M. A. (2014a). Industry platforms and ecosystem innovation. *Journal of Product Innovation Management*, *31*(3), 417–433.

Gawer, A., & Cusumano, M. A. (2014b). Industry Platforms and Ecosystem Innovation: Platforms and Innovation. *Journal of Product Innovation Management*, *31*(3), 417–433.

GDC. (2019, January 23). *Nearly 50% of devs support unionization, per new GDC State of the Industry report*. GDC. https://gdconf.com/news/nearly-50-devs-support-unionization-new-gdc-state-industry-report

Geeter, D. (2019, February 26). *Twitch created a business around watching video games — here's how Amazon has changed the service since buying it in 2014*. CNBC; CNBC.

https://www.cnbc.com/2019/02/26/history-of-twitch-gaming-livestreaming-and-youtube.html

Ghazawneh, A., & Henfridsson, O. (2010). Governing third-party development through platform boundary resources. *The International Conference on Information Systems (ICIS)*, 1–18.

Gittleson, K. (2014, August 25). Amazon buys game site Twitch. *BBC*. https://www.bbc.com/news/technology-28930781

Gough, C. (2018, June 1). *Global video games market value 2021 | Statista*. Statista. https://www.statista.com/statistics/246888/value-of-the-global-video-game-market/

Gregory Mankiw, N. (2016). *Principles of Economics*. Cengage Learning.

Gros, D., Wanner, B., Hackenholt, A., Zawadzki, P., & Knautz, K. (2017). World of Streaming. Motivation and Gratification on Twitch. *Social Computing and Social Media. Human Behavior*, 44–57.

Guide. (2020, April 1). Twitch Developers. https://dev.twitch.tv/docs/api/guide

Guzzle, PHP HTTP client — Guzzle Documentation. (n.d.). Retrieved April 2, 2020, from http://docs.guzzlephp.org/en/stable/

Hackernoon. (2020, January 17). *The Gaming Ecosystem Explained*. Hackernoon. https://hackernoon.com/the-gaming-ecosystem-explained-nk1d32ts

Hazée, S., Delcourt, C., & Van Vaerenbergh, Y. (2017). Burdens of Access: Understanding Customer Barriers and Barrier-Attenuating Practices in Access-Based Services. *Journal of Service Research*, *20*(4), 441–456.

Heaney, D., & Hamilton, I. (2020, April 1). *Steam Hardware Survey Revamped, Will Be Reliable Estimate Of PC VR Headset Ownership*. UploadVR. https://uploadvr.com/steam-hardware-survey-revamp/

Heise Medien. (2020, May 10). *heise online Preisvergleich*. Heise Medien. https://www.heise.de/preisvergleich/g-skill-aegis-dimm-kit-16gb-f4-3000c16d-16gisb-a1474853.html?hloc=at&hloc=en

Hicks, J. R., & Allen, R. G. D. (1934). A Reconsideration of the Theory of Value. Part I. *Economica*, *1*(1), 52–76.

Hitt, K. (2019, December 27). *The Top 10 Esports of 2019 by Total Prize Pool | The Esports Observer | home of essential esports business news and insights*. The Esports Observer | home of Essential Esports Business News and Insights. https://esportsobserver.com/biggest-esports-2019-prize-pool/

Hoelck, K., & Ballon, P. (2015). *Competitive Dynamics in the ICT Sector: Strategic Decisions in Platform Ecosystems*. https://papers.ssrn.com/abstract=2763648

Hoelck, K., Cremer, S., & Ballon, P. (2016). *Cross-Platform Effects: Towards a Measure for Platform Integration Benefit*. https://papers.ssrn.com/abstract=2817994

Horvat, F. (n.d.). *laravel-eloquent-join*. Github. Retrieved April 2, 2020, from https://github.com/fico7489/laravel-eloquent-join

IGN. (2016, October 13). *How to Use Twitch with PS4 - PlayStation 4 Wiki Guide - IGN*. https://www.ign.com/wikis/playstation-4/How_to_Use_Twitch_with_PS4

Influencer Marketing Hub. (2018, July 4). *How to Make Money on Twitch [Updated Feb 2020]*. Influencer Marketing Hub. https://influencermarketinghub.com/make-money-on-twitch/

Juul, J. (2009). Fear of failing? the many meanings of difficulty in video games. *The Video Game Theory Reader*, *2*(237-252). https://www.academia.edu/download/51135850/Juul__Fear_of_Failing_Video_Games.pdf

Karhu, K., Gustafsson, R., & Lyytinen, K. (2018). Exploiting and Defending Open Digital Platforms with Boundary Resources: Android's Five Platform Forks. *Information Systems Research*, *29*(2), 479–497.

Katz, M. L., & Shapiro, C. (1985). Network Externalities, Competition, and Compatibility. *The American Economic Review*, *75*(3), 424–440.

Khomh, F., Dhaliwal, T., Zou, Y., & Adams, B. (2012). Do faster releases improve software quality?

An empirical case study of Mozilla Firefox. *2012 9th IEEE Working Conference on Mining Software Repositories (MSR)*, 179–188.

Kuchera, B. (2019, January 24). *Does Valve deserve Steam's 30 percent cut? Many developers say no*. Polygon; Polygon. https://www.polygon.com/2019/1/24/18196154/steam-developers-revenue-epic-games-store

Leftronic. (2019, November 20). *25+ Incredible Twitch Statistics to Know in 2020*. Leftronic. https://leftronic.com/twitch-statistics/

Levitt, T. (1965, November 1). Exploit the Product Life Cycle. *Harvard Business Review*. https://hbr.org/1965/11/exploit-the-product-life-cycle

Lindblom, C. E. (1959). The science of" muddling through." *Public Administration Review*, 79–88.

Maxon Computer. (2020). *Cinebench R20*. MAXON | 3D FOR THE REAL WORLD. https://www.maxon.net/en-us/products/cinebench-r20-overview/

Muffatto, M., & Roveda, M. (2000). Developing product platforms:: analysis of the development process. *Technovation*, *20*(11), 617–630.

Munzert, S., Rubba, C., Meißner, P., & Nyhuis, D. (2014). *Automated Data Collection with R: A Practical Guide to Web Scraping and Text Mining*. John Wiley & Sons.

Nalebuff, B. J., Brandenburger, A., & Maulana, A. (1996). *Co-opetition*. HarperCollinsBusiness London.

Nambisan, S., & Sawhney, M. (2011). Orchestration Processes in Network-Centric Innovation: Evidence From the Field. *Academy of Management Perspectives*, *25*(3), 40–57.

Netflix. (2019). *Letter to Shareholders January 2019*. Netflix. https://s22.q4cdn.com/959853165/files/doc_financials/quarterly_reports/2018/q4/01/FINAL-Q4-18-Shareholder-Letter.pdf

Onyett, C. (2011, January 7). *Valve's Next Game - IGN*. IGN. https://web.archive.org/web/20121109072424/http://www.ign.com/articles/2011/01/07/valves-n ext-game

Otwell, T. (2016). *Laravel Introduction*. https://laravel.com/docs/4.2/introduction

Park, K., Seamans, R., & Zhu, F. (2018). *Multi-Homing and Platform Strategies: Historical Evidence from the US Newspaper Industry*. https://doi.org/10.2139/ssrn.3048348

Pindyck, R. S., & Rubinfeld, D. L. (2013). *Microeconomics*. Pearson.

Ram, S., & Sheth, J. N. (1989). Consumer Resistance to Innovations: The Marketing Problem and its solutions. In *Journal of Consumer Marketing* (Vol. 6, Issue 2, pp. 5–14). https://doi.org/10.1108/eum0000000002542

Razali, N. M., Wah, Y. B., & Others. (2011). Power comparisons of shapiro-wilk, kolmogorov-smirnov, lilliefors and anderson-darling tests. *Journal of Statistical Modeling and Analytics*, *2*(1), 21–33.

Rietveld, J., Schilling, M. A., & Bellavitis, C. (2019). Platform Strategy: Managing Ecosystem Value Through Selective Promotion of Complements. *Organization Science*, *30*(6), 1232–1251.

Rochet, J. C., & Tirole, J. (2003). Platform competition in two-sided markets. *Journal of the European Economic Association*. https://academic.oup.com/jeea/article-abstract/1/4/990/2280902

Schilling, M. A. (2003). Technological Leapfrogging: Lessons from the U.S. Video Game Console Industry. *California Management Review*, *45*(3), 6–32.

Shapiro, Shapiro, C., Carl, S., & Varian, H. R. (1998). *Information Rules: A Strategic Guide to the Network Economy*. Harvard Business Press.

Sherwin, N. (2019, September 24). *Do Streaming Metrics on Twitch Affect Game Sales?* Medium; Towards Data Science. https://towardsdatascience.com/do-streaming-metrics-on-twitch-affect-game-sales-cbb4e0ee90e0

Sjöblom, M., & Hamari, J. (2017). Why do people watch others play video games? An empirical study on the motivations of Twitch users. *Computers in Human Behavior*, *75*, 985–996.

Smith, C. (52 C.E.). *amazing Twitch stats and facts*. https://videogamesstats.com/twitch-stats-facts/

Sony. (2020, January 7). *PLAYSTATION™NETWORK MONTHLY ACTIVE USERS REACHES 103 MILLION*. Sony Interactive Entertainment. https://www.sie.com/en/corporate/release/2020/200107.html

Statista. (2016, August 25). *Gaming companies view of customers' favorite gaming platforms 2016 | Statista*. Statista. https://www.statista.com/statistics/608933/gaming-companies-customer-preferred-gaming-platforms-worldwide/

Statista. (2018a). *Steam - gaming platform*. Statista. https://www.statista.com/study/51846/steam-gaming-platform/

Statista. (2018b). *Video games market in the U.S*. Statista. https://www.statista.com/study/12321/video-game-market-in-the-united-states-statista-dossier/

Statista. (2020, April). *Global eSports viewership by viewer type 2022 | Statista*. Statista. https://www.statista.com/statistics/490480/global-esports-audience-size-viewer-type/

Sturgeon, P. (2018, May 16). *What is API Rate Limiting All About?* https://apisyouwonthate.com/blog/what-is-api-rate-limiting-all-about

Svahn, F., & Henfridsson, O. (2012). The Dual Regimes of Digital Innovation Management. *2012 45th Hawaii International Conference on System Sciences*, 3347–3356.

Takahashi, D. (2018, February 21). *Newzoo: Esports could hit 380 million fans and $906 million in revenues in 2018*. VentureBeat; VentureBeat. https://venturebeat.com/2018/02/21/newzoo-esports-could-hit-380-million-fans-and-906-million-in-revenues-in-2018/

Taylor, M., & Taylor, A. (2012). The technology life cycle: Conceptualization and managerial implications. *International Journal of Production Economics*, *140*(1), 541–553.

Twitch. (2020a). *TwitchCon*. TwitchCon. https://www.twitchcon.com/

Twitch. (2020b, May 9). *Twitch Prime*. Twitch; Twitch. https://twitch.amazon.com/tp

Twitch. (2020c, May 9). *Twitch Turbo Guide*. Twitch Turbo Guide; Twitch. https://help.twitch.tv/s/article/twitch-turbo-guide?language=en_US

Valentine, R. (2019, December 19). *Mixer, Facebook Gaming chipped away at Twitch market share in 2019*. GamesIndustry.biz. https://www.gamesindustry.biz/articles/2019-12-19-mixer-facebook-gaming-chipped-away-at-twitch-market-share-in-2019

Valve. (2020a). *Steam-community market*. Steam Community Market. https://steamcommunity.com/market

Valve. (2020b, March 29). *About Us - Valve Corporation*. Valve Corporation. https://www.valvesoftware.com/da/about

Valve Corporation. (2010, July). *Steam Web API Terms of Use*. https://steamcommunity.com/dev/apiterms

Wang, H., & Sun, C.-T. (2011). Game reward systems: Gaming experiences and social meanings. *DiGRA Conference*, *114*. http://dx.doi.org/

Wan, Y., & Yang, X. (2019). An empirical study of the self-fulfilling prophecy effect in Chinese stock market. *The Journal of Finance and Data Science*, *5*(2), 116–125.

Weidner, D. (2020). *laravel-goutte*. Github. https://github.com/dweidner/laravel-goutte

Wessel, M., Thies, F., & Benlian, A. (2017). Competitive Positioning of Complementors on Digital Platforms: Evidence from the Sharing Economy. *ICIS 2017 Proceedings*. https://aisel.aisnet.org/icis2017/Peer-to-Peer/Presentations/21/

Wessel, M., Thies, F., Benlian, A., & Others. (2017). Competitive Positioning of Complementors on Digital Platforms: Evidence from the Sharing Economy. *ICIS*. https://www.researchgate.net/profile/Michael_Wessel/publication/320299940_Competitive_Po sitioning_of_Complementors_on_Digital_Platforms_Evidence_from_the_Sharing_Economy/links/5a7028c1a6fdcc33daa83464/Competitive-Positioning-of-Complementors-on-Digital-Platforms-Evidence-from-the-Sharing-Economy.pdf

Williamson, O. E. (1975). *Markets and hierarchies, analysis and antitrust implications: a study in the economics of internal organization*. Free Press.

Yalcin, T., Ofek, E., Koenigsberg, O., & Biyalogorsky, E. (2013). Complementary Goods: Creating, Capturing, and Competing for Value. *Marketing Science*, *32*(4), 554–569.

Yan, R., & Bandyopadhyay, S. (2011). The profit benefits of bundle pricing of complementary products. *Journal of Retailing and Consumer Services*, *18*(4), 355–361.

Yoo, Y., Henfridsson, O., & Lyytinen, K. (2010). Research Commentary—The New Organizing Logic of Digital Innovation: An Agenda for Information Systems Research. *Information Systems Research*, *21*(4), 724–735.

Zagal, J. P., Nussbaum, M., & Rosas, R. (2000). A Model to Support the Design of Multiplayer Games. *Presence: Teleoperators and Virtual Environments*, *9*(5), 448–462.

Zipp, R. (2020). *Laravel-Twitch*. Github. https://github.com/romanzipp/Laravel-Twitch

Appendix

1. Data set

We provide the data set as a compressed SQL-file under the following links. To ensure the accessibility of the data set also outside of the CBS network we provide it additionally on Google Drive.

Internal CBS network	https://studentcbs-my.sharepoint.com/:u:/g/personal/leze18ab_student_cbs_dk/EWpY1HF1NmxButavGUIoj7IB-cdVmT1i5oj7AF7xfEbFJw?e=XQ9er4
Google Drive	https://drive.google.com/open?id=1xMF41z850ApMORMI_045xm0zAHQSQ65z

2. Power BI project

Furthermore, to also ensure transparency with regard to our data analysis, we provide the executable Power BI project with corresponding R-scripts on CBS SharePoint and Google Drive.

R-scripts used in the paper can be found within the Power BI file if not linked explicitly in the paper.

Internal CBS network	https://studentcbs-my.sharepoint.com/:u:/g/personal/leze18ab_student_cbs_dk/EcIeWN3AUX5CryFw6JZ2Bv8BETXdR2YAqkRWRZXpRbYQFw?e=KpTTP6
Google Drive	https://drive.google.com/drive/folders/1dmc4Oxk7YXIJ7_LB4boTwMdyT8-Dypeb?usp=sharing

3. Data collection application

Finally the source code for the data collection is provided through Gitlab.com, Google Drive and CBS sharepoint.

Gitlab.com	https://gitlab.com/cbs-master
Internal CBS network	https://studentcbs-my.sharepoint.com/:u:/g/personal/leze18ab_student_cbs_dk/EWnrIAoTFN5HnRx8Gxv_IzoBxk_QlNcS217G-SdVP3LiqQ?e=1uPH4j https://studentcbs-my.sharepoint.com/:u:/g/personal/leze18ab_student_cbs_dk/ERBPrIUvr6BAq7HhlyFalocBli29KXuGjjU9b6KgNVVMpQ?e=E69xcx
Google Drive	https://drive.google.com/open?id=1_nOxGYeSzyXQOEC9imb2-WahTR99_QMx https://drive.google.com/open?id=1dgnvey-hLEnxa0Ehlg5HBlYI1NWSyLaI