



Looking into the past: Audience heterogeneity and the inconsistency of market signals

Strategic Organization

1–33

© The Author(s) 2022



Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/14761270221139760

journals.sagepub.com/home/soq**Pengfei Wang** 

BI Norwegian Business School, Norway

Abstract

Prior literature has emphasized that inconsistency of market signals leads to evaluation penalty. However, limited attention has been paid to the heterogeneity of audiences who deal with inconsistency. I argue that audiences differ in the extent to which they process different market signals, which may largely shape the effect of signal inconsistency. When audiences fail to process all signals, they may not perceive signal inconsistency, thereby weakening its effect on product evaluation. It is hence important to investigate audience heterogeneity in theorizing signal inconsistency. In this study, I focus on the distinction between two important audience groups: professional critics and end consumers. Specifically, I argue that signal inconsistency exerts a stronger effect on critics' evaluations than on consumers' evaluations, because critics are more likely than consumers to process various market signals. I argue further that critics can act as an important intermediary to bridge the effect of signal inconsistency on consumers, even though consumers may not process all signals themselves. I test these ideas in a sample of video games released between 2001 and 2016 and find general support.

Keywords

categories, regression analysis, reputation, research methods, signaling theory, status, topics and perspectives

Introduction

Market signals play an important role in shaping evaluation as they help reduce audiences' uncertainty about quality (Spence, 1974). Research on the signaling process has proliferated over the last few decades, identifying various positive and negative signals in a wide array of contexts (Connelly et al., 2011; Steigenberger and Wilhelm, 2018). While the literature has well established how and when one particular market signal is effective (Kim and Jensen, 2014; Negro et al., 2015; Podolny, 1993; Pollock and Gulati, 2007), recent studies emphasize that most firms release multiple signals that convey quality information, and draw attention to an important issue of inconsistency when the signals are incongruent (Wang and Jensen, 2019). Signal inconsistency can be harmful to firms

Corresponding author:

Pengfei Wang, Department of Strategy and Entrepreneurship, BI Norwegian Business School, Nydalsveien 37, Oslo 0484, Norway.

Email: [Pengfei.wang@bi.no](mailto: Pengfei.wang@bi.no)

who send those signals, because it creates ambiguity about their quality, raising audiences' concerns about what quality to expect from the firms (Zhao and Zhou, 2011).

However, for signal inconsistency to influence evaluation, a key scope condition is required that audiences process all market signals of a firm (Connelly et al., 2011; Madsen and Rodgers, 2015). Only when audience members pay attention to various signals, could they perceive any inconsistency of the signals. Failing to process all signals may lead audiences to misperceive the level of inconsistency, thereby weakening the adverse effect of signal inconsistency on their evaluation.¹ However, this critical issue has been underexplored in the research of signal inconsistency. Indeed, prior literature on signal inconsistency has mostly examined its impact on one specific type of audience evaluation (Jensen and Wang, 2018; Vergne et al., 2018), with limited discussion on the heterogeneity of audiences who deal with the signals. This neglect is a bit surprising, given the emphasis on audience heterogeneity in recent literature on social evaluation (Durand and Haans, 2022; Falchetti et al., 2022; Sharkey et al., 2022). Kim and Jensen (2011), for instance, show that opera critics and fans react differently to market identity; Pontikes (2012) finds that investors and consumers have different preferences in evaluating startups. Bringing audience heterogeneity into research on signal inconsistency, I emphasize that audiences differ in their processing scope of various signals, so that they may not react to signal inconsistency in the same manner. By doing so, I shift attention from the general signal inconsistency discount (Vergne et al., 2018) to how the evaluative concern of signal inconsistency differs across audience groups.

I focus specifically on two types of market audiences—professional critics and end consumers. Critics and consumers are important audiences in the product market, as their evaluations are consequential for firms' economic exchanges (Kim and Jensen, 2011). Whereas consumers determine more directly firms' market performance (Zhu and Zhang, 2010), critics influence the demand for goods in an indirect way (Reinstein and Snyder, 2005; Zuckerman, 1999). It seems intuitive that both critics and consumers are susceptible to signal inconsistency, as they rely on market signals to form evaluative judgments. I emphasize, however, that the two types of audiences differ in terms of their attention to various signals (Kovács and Hannan, 2010), a key scope condition for signal inconsistency. While critics often analyze a firm exhaustively in their evaluation process (Hsu, 2006a; Sharkey et al., 2022), consumers rarely process all the signals of the firm. Since signaling can only be effective to the extent that audiences receive and interpret signals (Madsen and Rodgers, 2015), consumers would be less likely than critics to observe the inconsistency of the signals. As such, I argue that the impact of signal inconsistency is weaker on consumers than critics.

Second, on top of the direct effects, I stress that critics can act as an important market mediation to bridge the effect of signal inconsistency on consumers. While prior literature has emphasized that different audience groups can react differently to market practices and signals (Pontikes, 2012), the mutual influence between audience groups is much less discussed in studies of audience heterogeneity. Different groups of audiences are not completely isolated, but commonly affect each other's evaluations (Kovács and Sharkey, 2014). For instance, while peers in movie production have different evaluative focus than professional critics (Cattani et al., 2014), winning Oscars from *peer* voting might also influence how a film is evaluated by *critics* after the ceremony. As such, accounting for the link between different groups is both necessary to estimate each audience group's *direct* (instead of *total*) reaction to signal inconsistency, and important to understand the *pathways* through which signal inconsistency affects the evaluations of different audiences (Kovács and Sharkey, 2014). The link is particularly pronounced between critics and consumers, since critics have long been known for shaping the attention and evaluation of consumers (Basuroy et al., 2020; Blank, 2007), albeit critics may sometimes conform to the taste of consumers as well (Pang et al., 2022). By emphasizing the interaction between critics and consumers, I argue that even

though consumers may not have a full coverage of a firm's signals and hence be less reactive to its signal inconsistency, critics can mediate the effect of signal inconsistency on consumers.

By specifically examining the temporal incongruence of reputation signals, this article also emphasizes a new form of inconsistency. Prior studies have focused on two types of inconsistency. Some scholars highlight *dimensional* inconsistency, the extent to which signals are different across dimensions (Vergne et al., 2018; Zhao and Zhou, 2011); others pay attention to *categorical* inconsistency, the extent to which signals are incongruent spanning market categories (Jensen and Wang, 2018). In this article, I shift attention to *temporal* inconsistency of a firm's quality signals over time, which remains underexplored. Whereas the literature on reputation has mostly examined the average level of past quality (Shapiro, 1983), this study emphasizes that the variation in quality levels in the past also imposes substantial constraints on new product evaluation. That is, audiences are sensitive to the consistency of reputation signals, above and beyond the average level of the signals.

I test these hypotheses using a sample of video games. This empirical setting is appropriate as it allows me to observe separately critic and consumer evaluations of games on Metacritic (Chen et al., 2012). I first develop a framework to examine and compare the effects of signal inconsistency on critic and consumer evaluations. Empirically, signal inconsistency of a game developer is defined as the extent to which the quality levels of its past games are incongruent. And then I employ structural equation modeling to formally disentangle the direct effect of signal inconsistency on consumers from its indirect effect due to the mediating role of critics. The results show that while the *direct* effects of signal inconsistency on critic and consumer evaluations are both negative, the former's effect size is more than two times as large as the latter's. I also find support for a mediating effect of critic evaluations on the relationship between signal inconsistency and consumer evaluations. This explains why the *total* effects of signal inconsistency on critic and consumer evaluations are comparable, but the two *direct* effects are substantially different.

Theory and hypotheses

Market signals and signal inconsistency

Uncertainty about quality is a widespread and important feature of markets, as quality information is commonly asymmetric between firms and their audiences (Shapiro, 1983; Spence, 1974). Firms who are insiders often obtain information that is not available to outside audiences. This is particularly so when products involve more experience attributes (e.g. movies, wines, and video games), such that one can only evaluate products through consumption (Nelson, 1970). To deal with the uncertainty, audiences tend to rely on various market signals of a firm and its products during quality evaluation (Ebbers and Wijnberg, 2010; Ni and Tang, 2013). Market signals are strategic measures that convey quality information about a firm or a product, influencing how audiences evaluate them in uncertain situations (Vergne et al., 2018). A positive signal enhances audiences' beliefs about a product's quality, whereas negative signals cause more concerns.

To evaluate a product, audiences may refer to both the signals that are specific to the product and the general signals of its producer. Audiences are sensitive to a variety of *product-level* signals (Connelly et al., 2011). High price and unconditional warranty of a product, for instance, make potential buyers to perceive good quality (Milgrom and Roberts, 1986); product recalls, on the contrary, contain negative information about the products and their producers (Rhee and Haunschild, 2006). At the same time, *producer-level* signals such as status and reputation constitute producers' vertical market identity that entails normative expectations, signaling the quality of their offerings (Podolny, 1993; Sorenson, 2014). Audiences usually believe products to be of high quality if they

are offered by firms who are of high status or high reputation (Obloj and Capron, 2011; Shapiro, 1983; Stern et al., 2014).

While earlier literature focuses on whether and when one particular signal is effective, recent studies shift attention to the inconsistency of multiple signals from the same source (Vergne et al., 2018). A firm often releases different signals that convey quality information (Ni and Tang, 2013), which can be either consistent or inconsistent. When multiple signals are congruent, they can cross-validate each other and enhance the effectiveness of the signaling process (Stern et al., 2014). Consistency among multiple signals is hence crucial in constructing a clear and robust market identity of firms. In reality, however, signals are often incongruent, leading to an issue of signal inconsistency. Signal inconsistency refers to the extent to which multiple signals convey incongruent quality information about the same firm (Connelly et al., 2011). Signal inconsistency can be detrimental to firms, because it creates ambiguity about their quality claims. That is, a firm's incongruent signals will create an interpretation challenge for external audiences, who cannot make sense of what quality to expect from the firm.

Signal inconsistency may exist in several forms. First, signal inconsistency occurs when signals about a product or a firm are inconsistent across different *dimensions* (Vergne et al., 2018). A product that is priced high but provides no warranty and a firm that engages in philanthropy but overcompensates its CEO are examples. When signals are inconsistent across dimensions, their positive claims are undermined, leading to devaluation in the market. Zhao and Zhou (2011), for instance, examine signal inconsistency of Californian wines along three dimensions (i.e. professional critics, wine appellations, and extra designations), and find evidence that wine prices are constrained by such inconsistency. Second, signal inconsistency also exists when a firm releases incongruent signals in different *categories* (Jensen and Wang, 2018). When a firm spans multiple market categories and sets very different prices for its products (e.g. luxury handbags and budget watches), for instance, audiences receive incongruent quality signals. Such signal inconsistency poses problems to a multiproduct firm, as audiences will likely find it challenging to make sense of the firm's overall quality. Third, signal inconsistency also occurs when a firm sends inconsistent signals over different *time periods*. A signal is essentially a snapshot pointing to unobservable quality at a particular point in time (Davila et al., 2003). Audiences will hence receive multiple product signals about one firm over time, which may be inconsistent from each other. For instance, rankings that signal the quality of universities may fluctuate across years (Askin and Bothner, 2016), and online ratings that are essential signals of sellers' quality may swing over time (Parker et al., 2017), both leading to the third form of signal inconsistency. In this article, I focus specifically on the third form of signal inconsistency (i.e. the temporal incongruence of reputation signals), though I expect that my theory can be applied to the other forms as well.

Past quality as market signals

While scholars have identified a variety of signals, one of the most visible signals is product quality in the past (Shapiro, 1983). That is because audiences tend to form an expectation of future quality that is directly based on past demonstrations (Jensen and Roy, 2008). When a firm's past offerings are of good quality, for instance, it gains a high reputation that will signal its ability to offer high-quality products or services in the future. Reputation will lead audiences to form more favorable expectations and evaluations (Jensen et al., 2012). Sports teams, for instance, pay more for players who have shown better performance in the past (Ertug and Castellucci, 2013); public firms choose auditors who have already demonstrated good technical skills (Jensen and Roy, 2008); buyers offer greater price premium in online transactions to sellers who have received more positive ratings (Obloj and Capron, 2011).

While past quality or reputation has been well established as a quality signal, it is not necessarily consistent (Parker et al., 2017). Players' performance, for instance, may fluctuate over seasons; firms' performance can go up and down; sellers might receive divergent online reviews for their products. The inconsistency of past quality is consequential. It challenges the effectiveness of past quality as a market signal. When a firm's past quality is stable and consistent, it can serve a valid and reliable signal based on which audiences form quality expectations and judgments (Shapiro, 1983). However, when a firm has delivered outputs of divergent quality, the function of its past quality as a signal becomes obscured (Connelly et al., 2011). Incongruent past quality, in other words, makes a firm's quality claims ambiguous (Parker et al., 2017). Being concerned with ambiguity, audiences may ignore such a firm, avoid transactions with it, and lower their evaluations of the firm (Vergne et al., 2018).

Signal inconsistency in the video games context

In the research context of video games, audiences face uncertainty about evaluating the quality of new games. To deal with the uncertainty, audiences usually consider the quality of a developer's previous games as an important market signal. The quality level of past games largely reflects a developer's reputation, as it indicates how its games perform relative to others in terms of perceived quality (Rietveld and Eggers, 2018). A game developer gains a good reputation from its past provision of high-quality games (Choi et al., 2018), which is commonly shown in various online platforms that constantly evaluate, rate, and rank games (Parker et al., 2017). Because of the signaling role of reputation (Obloj and Capron, 2011; Shapiro, 1983), a developer's new games will be more appealing if its past games have accumulated a higher level of average quality reputation.

While all developers hope for their games to gain a high reputation, even reputable developers (e.g. Ubisoft) cannot guarantee that their games are always at high-quality levels. When a developer's past games have displayed inconsistent quality, it runs into the problem of reputation signal inconsistency.² Audiences will find its quality claims ambiguous and cannot make sense of what quality to expect from such a developer (Wang and Jensen, 2019). Being concerned about the quality ambiguity induced by signal inconsistency, market audiences may devalue the developer and its new offerings (Hsu, 2006b; Zuckerman, 1999). That is, I expect that audience evaluation on a new game is contingent on the extent to which a developer's previous games emit congruent quality reputation signals, above and beyond the average level of the signals. Audiences will likely make negative evaluations of games from developers whose past game reputations are more inconsistent.³ I therefore formulate the baseline hypothesis as follows:

Hypothesis 1 (H1). The more inconsistent a developer's past reputation signals are, the lower its new game will be evaluated by audiences (both critics and consumers).

Signal inconsistency and audiences

Signaling is effective to the extent that audiences process the signals released by firms (Madsen and Rodgers, 2015). An emitted signal will have no effect if audiences do not look for the signal, or do not know what to look for and how to interpret (Connelly et al., 2011). As such, the impact of signal inconsistency will depend on the extent to which audiences observe and interpret all signals about an entity. For example, for the inconsistency between wine appellations and extra designations to take effect (Zhao and Zhou, 2011), it is necessary for consumers to understand wines' positions in both dimensions; for the inconsistency between corporate philanthropy and CEO over-compensation to influence evaluation (Vergne et al., 2018), it is vital for media to cover a firm's practices in both areas.

If audiences fail to process all of the relevant signals, they may end up with a biased perception about the level of signal inconsistency. For instance, if media focus on a firm's philanthropy practice but overlook its CEO overcompensation, they may not perceive any signal inconsistency; if consumers have little knowledge about wine appellations and extra designations, they may not sense a wine's signal inconsistency. While existing studies on cognition heuristics (Goldstein and Gigerenzer, 2002) would suggest that audiences do not have to possess perfect knowledge about market signals for signal inconsistency to remain important, it is still compelling that the effect of signal inconsistency is contingent on the extent to which audiences process various signals (Connelly et al., 2011).

However, in analyzing the role of signal inconsistency, prior studies often assume (implicitly) that various market signals are well processed by audiences (Jensen and Wang, 2018; Zhao and Zhou, 2011). Relaxing this assumption, I emphasize that there exists substantial heterogeneity between audience groups. Compared with end consumers, professional critics (or other market intermediaries) are more motivated to attend to the details of signals and process them systematically, in striving for accuracy in evaluations. A fundamental goal of critics is to offer principles for the ranking of products relative to one another in terms of perceived quality, which can establish themselves as an expert and guide market evaluation (Becker, 2008; Blank, 2007; Hsu, 2006a). Critics are hence motivated to help reveal hidden product quality and specify particular criteria for product assessment (Sharkey et al., 2022).

As a result, critics usually pay much closer attention to all the offerings and signals of an entity, to provide more comprehensive, detailed, and convincing evaluations. Selling-side analysts, for example, will investigate carefully all segments of a listed firm when publishing their forecasts (Zuckerman, 1999); movie critics commonly analyze the historical records of a film's directors and actors when posting their reviews; media journalists try to cover many relevant conducts of a firm when writing their articles (Vergne et al., 2018); video game critics acquire in-depth knowledge about a developer's past offerings before providing their evaluations on its newly released game (Zhao et al., 2018). As critics vigilantly scan the environment to process all relevant signals, they are more likely to sense the degree of signal inconsistency. As such, critics will be susceptible to signal inconsistency as expected.

While it is possible that a few end consumers have equivalent coverage as critics, the majority of consumers are less likely to process a complete set of signals (Kovács and Hannan, 2010). As compared with critics whose career reputation hinges on their evaluations, consumers are less motivated to search for all relevant quality signals in their own evaluation process. Missing some signals will then lead the consumers to overlook signal inconsistency, thereby weakening their reactivity to it. Put differently, in product evaluations, critics and consumers may consider a different set of information cues (Falchetti et al., 2022). While consumers are more likely to focus only on the product-level attributes, critics tend to make a more holistic evaluation of the product, its producer, and even the social and competition environment (Blank, 2007). More specifically, in evaluating a new game, critics are more likely to process the developer's reputation signals by considering the past demonstrations of its offerings in the market, so that they are better able to discern its degree of temporal signal inconsistency. Consumers, however, may not conduct a comprehensive analysis of the developer's past games, and hence fail to spot the inconsistency of reputation signals at the developer level. Taken together, I conjecture that end consumers are less attentive than professional critics to the inconsistency of reputation signals. Formally,

Hypothesis 1 (H2). The negative effect of signal inconsistency on evaluations is weaker for consumers than for critics.

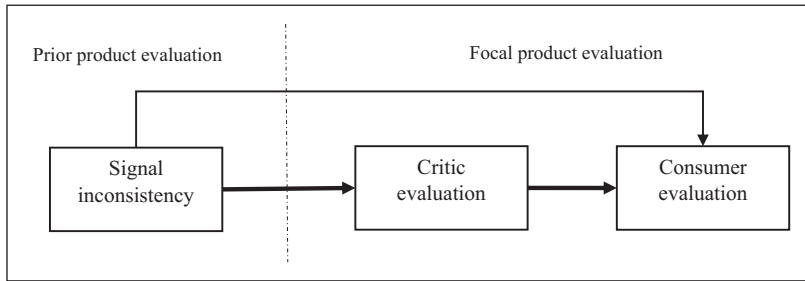


Figure 1. Conceptual framework.

However, critics may pass over their concerns about signal inconsistency to customers, through their substantial influences in the market. Because the evaluation of cultural or experiential products is complicated and difficult (Nelson, 1970), consumers often refer to critics in making their own assessments in various settings such as video games and movies (Zhao et al., 2018; Zuckerman and Kim, 2003). Indeed, being aided by communication technologies (e.g. Internet), critic evaluations have become widely available to consumers (Chen et al., 2012; Zhu and Zhang, 2010). Their evaluations play important roles in facilitating consumers' judgment, shaping consumers' theory of values,⁴ and drawing consumers' attention toward some criteria and away from the others (Durand and Boulongne, 2017). By doing so, critics directly or indirectly establish a particular schema for value assessment, which aims to guide the evaluation and taste of others in the market (Hsu, 2006a; Hsu et al., 2012b).

When consumers rely on critics' evaluations as an anchor to form assessments (d'Astous and Touil, 1999), they implicitly become susceptible to signal inconsistency.⁵ As a result, even though game consumers may not be able to clearly observe developer-level signal inconsistency themselves, they are influenced by the inconsistency indirectly when they accept the evaluation schema and standard of game critics. By contrast, if consumers do not refer to critics in their evaluation process, they would be less likely to sense the incongruence of reputation signals, reducing the impact of signal inconsistency.⁶ I therefore argue that critic evaluations work as a mediator between signal inconsistency and consumer evaluations.

Hypothesis 3 (H3). Critic evaluations mediate the effect of signal inconsistency on consumer evaluations.

Figure 1 depicts my conceptual framework. Signal inconsistency of past game quality is expected to influence both critic and consumer evaluations. However, its effect on critics is stronger than that on consumers. And finally, the effect of signal inconsistency on consumer evaluations is mediated by critic evaluations.

Methods

I test my hypotheses using data on video games released in the global market between 2001 and 2016. Video games are a type of experiential products that involve uncertainty, as consumers cannot know games' quality before their purchase and consumption. Even after consuming a game, many consumers may still be uncertain about what ratings to give. As such, market signals play an important role in shaping audience evaluations in this context. Indeed, prior

research has shown that the market acceptance of video games is largely influenced by various quality signals such as price, online word of mouth (Zhu and Zhang, 2010), and firm reputation (Choi et al., 2018).

The video game market is appropriate to test signal inconsistency, because there is often quality variation of different games released from the same developers. I define signal inconsistency of a game developer as the extent to which its past games display incongruent quality levels (Parker et al., 2017). Figure 2 depicts the quality levels of games in my sample developed by Nintendo and Ubisoft from 2001 to 2016. While both developers' games show inconsistent quality, it seems that Nintendo games are less inconsistent than Ubisoft games. Indeed, as reported in the table, Ubisoft shows a larger signal inconsistency than Nintendo in the three periods, according to my measurement (explained below). See more examples in Appendix 1.

Second, this setting is also suitable to compare evaluations from different types of audiences. The game ratings of critics and consumers are separately recorded in widely used datasets, which allows me to explore how they may react differently to signal inconsistency. Specifically, I collect the key variables, critic and user evaluative ratings, from the review aggregator *Metacritic.com*. Metacritic keeps track of more than 300 game review publications, from which it creates Metascore that indicates the overall evaluations by critics (Rietveld, 2018). Game users also constantly post their ratings of games on Metacritic, which can be used as a proxy of consumer evaluations. Since games may be released either exclusively for one platform or for multiple platforms, the data are organized at the *game-platform* level, with a single observation per game for each platform on which it is released (Rietveld and Eggers, 2018). I collect other information (e.g. game sales, year of release, platform, and genre) from VGChartz.com. The final sample includes 4645 observations from 478 developers.⁷

Third, and more importantly, I choose to analyze video games because temporal sequence is relatively clear in this context, such that consumer evaluations usually come after critic evaluations. This is an important scope condition for the mediation hypothesis. While temporal sequence can be context-dependent, critics do provide evaluations prior to the majority of consumers in the video game ratings. Many critic ratings are published before official game releases, or very shortly after the games are released (Zhu and Zhang, 2010). It is common in the video game industry that many critics have the opportunity to play a game several days or weeks before its release date (Stuart, 2015). This is intuitive since critics aim to be the pioneers guiding the evaluations of consumers (Blank, 2007). Users, by contrast, mostly post their ratings in a window after critics (please see Appendix 2 for more discussions).⁸

Dependent variables

I refer to critic and user ratings of a new product as how it is evaluated by professional critics and end consumers, respectively (Hsu, 2006b). Specifically, I measure *critic evaluations* using the Metascores of games. Metascore is aggregated by Metacritic from game review publications (Rietveld and Eggers, 2018). Larger scores indicate more favorable evaluations by game critics, and vice versa. Because of its prevalence, Metascore is described as the video game industry's "premier" review aggregator (Leack, 2015). Indeed, it has been used by many companies (e.g. Microsoft) when making important decisions about game development (Remo, 2008). It ranges between 0 and 100, but I scale it by 10, to make it comparable with the range of consumer evaluations. Similarly, I measure *consumer evaluations* with average user ratings from Metacritic. When a game is well accepted by end consumers, one should expect higher user ratings. It can range from 0 to 10.

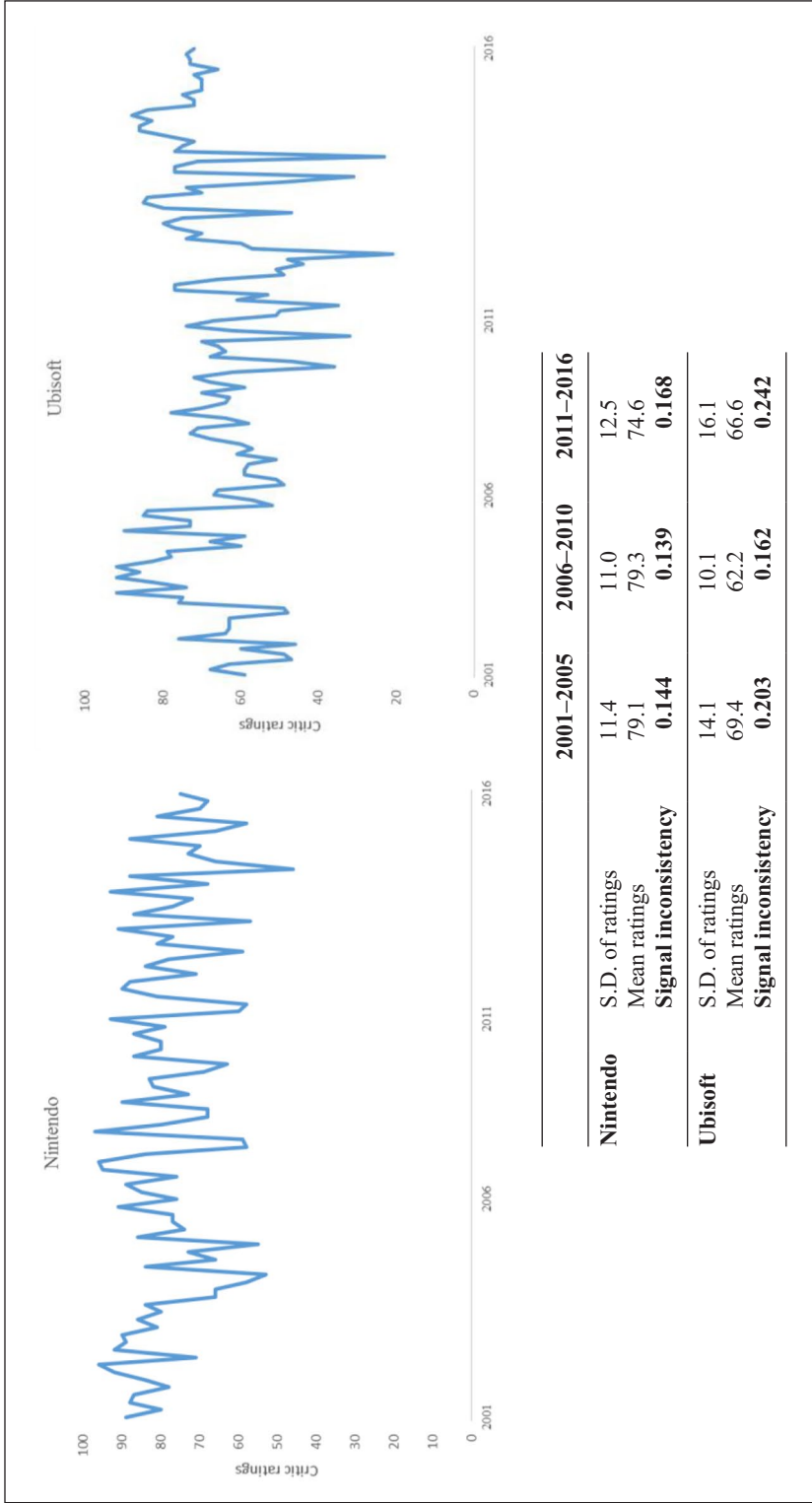


Figure 2. Critic evaluations of games by Nintendo and Ubisoft, 2001–2016.

Independent variable

Signal inconsistency. I measure signal inconsistency as the extent to which the quality reputation of a developer's past games is incongruent. In line with previous studies (Ebbers and Wijnberg, 2010; Parker et al., 2017), I use mean critic ratings to indicate a game's quality reputation, which is the most salient quality indicator of the game relative to others in the marketplace (in a robustness check below, I also use mean user ratings as an alternative measure of game quality reputation). Specifically, I calculate the signal inconsistency of a developer as the standard deviation of mean critic ratings of its games in the past 5 years before the focal game (e.g. a 5-year time window with 1-year time lag), scaled by the mean of past ratings

$$\text{Signal inconsistency}_{it} = \frac{\sqrt{\frac{1}{l-1} \sum_{j=1}^l (S_{ijt} - M_{it})^2}}{M_{it}}$$

$$M_{it} = \frac{\sum_{j=1}^l S_{ijt}}{l}$$

where S_{ijt} is the quality ratings (i.e. rating score) of game j by developer i between year $t-5$ and $t-1$, l is the number of games developed by i ; M_{it} is the mean of i 's game quality ratings between year $t-5$ and $t-1$. Scaling standard deviation by mean (i.e. the coefficient of variation) helps rule out the effect that is simply caused by the average ratings.

Control variables

Developer reputation. While focusing on the inconsistency of quality reputational signals, I control for the average level of them (i.e. developer reputation) that can also affect market acceptance (Choi et al., 2018). However, there are empirical challenges if simply adding the average as a control variable, because including the average variable twice on the right-hand side might lead to estimation bias. Inspired by related literature (Han and Pollock, 2021), I create a dummy variable that indicates whether a developer has a high reputation, to mitigate the potential concerns about collinearity. I rank all game developers based on the average reputation, and assign a value of 1 to a developer (i.e. high reputation) if it is within the top 20th percentile, and 0 otherwise.⁹

Developer status. The status of game developers can influence audience evaluation. Following previous studies, I assume an actor with high status if it wins top awards (Ertug and Castellucci, 2013). Specifically, I include a dummy variable indicating whether any games of a developer are awarded *Game Developers Choice Awards*. The awards are announced annually by the Game Developers Conference to the most innovative and creative game designers (De Vaan et al., 2015). The awards are introduced in 2001, including awards such as Game of the Year, Best Audio, Best Debut, Best Design, Innovation Award, Best Narrative, Best Technology, and Best Visual Art. A game developer is considered as of high status if any of its games win any of the awards in the past 5 years.

Number of genres. While focusing on reputational signals, I control for category spanning of game developers (i.e. grades of membership) (Hsu et al., 2009). Games are categorized into genres such as Strategy, Fighting, Role-playing, and Sports. I count the number of game genres that a developer has spanned in the past 5 years.

Genre fuzziness. Following (Hsu et al., 2012a), I also control for the extent to which a developer's genre participation is ambiguous. I first calculate a developer i 's grade of membership in a genre g , $\mu_i(g)$, based on the number of games it develops in the genre divided by the number of games it develops in the past 5 years. I then compute the fuzzy density for a genre, $N(g)$, as the sum of the grades of membership of developers that affiliate with the genre. Genre contrast, $N(g)/N$, is measured as the fuzzy density divided by the number of developers with at least one game in that genre over the past 5 years. Fuzziness is calculated as the opposite of contract

$$Fuzziness(g) = 1 - N(g)/N$$

To measure the genre fuzziness of a developer i , I sum up the product of a developer's grade of membership in a genre with the genre's fuzziness

$$Genre\ Fuzziness_{it} = \sum \mu_{it}(g) \times Fuzziness(g)$$

I control for game distribution patterns by including a dummy variable, *Multi-homing*, indicating if a game is launched for multiple platforms. A game's evaluation may also depend on the number of competing games that are available in the same year, genre, and platform. I therefore include a variable, *Niche Density*, to rule out that effect. I control for *Sales of publisher*, the global sales of game publishers in the past 5 years. I introduce four platform dummy variables: *Platform-PS*, *Platform-Xbox*, *Platform-Nintendo*, and *Platform-PC*, to indicate if a game is developed for any of those major platforms. The four platforms cover over 92% of all the games in my final sample.¹⁰ To account for any variances across time and categories, I include year fixed-effects, game genre fixed-effects, and Entertainment Software Rating Board (ESRB) rating fixed-effects.¹¹

Analysis and results

Following previous studies (Hsu et al., 2009), I employ Tobit regressions to test my Hypotheses.¹² To address the concern that games are not completely independent, I cluster standard errors by game developers and report robust standard errors (Singh and Fleming, 2010). To test the mediation effect of critic evaluations, I employ structural equation models (SEM) using the *sem* command in Stata (Zhao et al., 2018). SEM accounts for the fact that signal inconsistency may affect consumer evaluations in both direct way and indirect way through critic evaluations. It mitigates measurement errors that may bias the estimation of mediating effects (Shaver, 2005) and corrects error terms across systems of equations simultaneously to address potential endogeneity. To better identify the system, I have to specify one explanatory variable that directly affects critic evaluations but does not have a direct effect on consumer evaluations. Specifically, I include a dummy variable of *main genre* indicating whether a game falls into the main genre of the developer. I identify a genre as main genre if it includes the largest proportion of games that a developer has released in the past 5 years. The analysis suggests that it meets the empirical requirement as it has a significant effect on critic evaluations, but its direct effect on consumer evaluations is non-significant.

Table 1 shows descriptive analysis and correlations. It is worth noting that while the pairwise correlation between *critic evaluations* and *consumer evaluations* is positive ($r=0.54$), they are not completely coupled, suggesting potential evaluative disparity between the two types of

Table 1. Descriptive analysis and correlations.

Variables	Mean	SD	Min.	Max.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	
1 Consumer evaluations	7.16	1.41	0.50	9.60	1.00															
2 Critic evaluations	7.10	1.33	1.30	9.80	0.54	1.00														
3 Signal inconsistency	0.13	0.08	0.00	0.67	-0.11	-0.19	1.00													
4 Developer reputation	0.20	0.40	0.00	1.00	0.16	0.35	0.43	1.00												
5 Developer status	0.06	0.25	0.00	1.00	0.06	0.15	0.01	0.10	1.00											
6 No. of genres (ln)	1.31	0.50	0.69	2.48	0.01	-0.03	0.37	-0.28	0.13	1.00										
7 Niche density (ln)	2.92	0.78	0.69	4.53	0.05	-0.05	0.00	-0.03	0.01	0.00	1.00									
8 Genre fuzziness	0.41	0.07	0.13	0.63	-0.11	-0.16	0.11	-0.24	0.04	0.23	0.00	1.00								
9 Multi-homing	0.70	0.46	0.00	1.00	-0.12	0.03	-0.02	0.01	-0.03	-0.06	0.05	0.05	1.00							
10 Main genre of developer	0.58	0.49	0.00	1.00	0.02	0.11	-0.11	0.16	-0.06	-0.30	0.08	-0.08	0.12	1.00						
11 Sales of publisher (ln)	4.12	1.54	0.00	6.48	-0.04	0.17	-0.05	0.13	0.08	0.16	-0.01	0.17	0.13	-0.02	1.00					
12 Platform-PlayStation	0.41	0.49	0.00	1.00	0.06	-0.02	-0.02	0.01	-0.08	-0.05	0.12	-0.07	0.01	0.08	-0.11	1.00				
13 Platform-Xbox	0.24	0.43	0.00	1.00	-0.07	0.02	-0.01	0.04	0.01	-0.05	0.01	0.01	0.20	0.04	0.04	-0.47	1.00			
14 Platform-Nintendo	0.18	0.39	0.00	1.00	-0.03	-0.11	0.09	-0.15	0.01	0.17	0.06	0.21	-0.23	-0.13	0.08	-0.40	-0.27	1.00		
15 Platform-PC	0.08	0.28	0.00	1.00	0.06	0.14	-0.09	0.10	0.05	-0.12	-0.23	-0.04	0.02	0.06	0.01	-0.25	-0.17	-0.14	1.00	

SD: standard deviation.

Table 2. Tobit regressions of signal inconsistency on critic and consumer evaluations.

Variables	Critic evaluations		Consumer evaluations			
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Signal inconsistency		-1.651*** (0.425)		-1.780*** (0.510)		-0.710† (0.392)
Developer reputation	0.920*** (0.081)	0.805*** (0.082)	0.500*** (0.071)	0.377*** (0.074)	-0.099 (0.065)	-0.145* (0.068)
Developer status	0.633*** (0.103)	0.642*** (0.101)	0.271** (0.104)	0.281** (0.105)	-0.142 (0.087)	-0.136 (0.088)
No. of genres (ln)	0.204* (0.087)	0.272** (0.087)	0.143* (0.072)	0.216** (0.073)	0.010 (0.050)	0.040 (0.054)
Niche density (ln)	0.017 (0.038)	0.014 (0.037)	0.025 (0.044)	0.022 (0.043)	0.014 (0.032)	0.013 (0.032)
Genre fuzziness	-1.734* (0.749)	-1.606* (0.727)	-0.738 (0.731)	-0.600 (0.711)	0.392 (0.618)	0.441 (0.614)
Multi-homing	0.029 (0.065)	0.033 (0.065)	-0.173** (0.063)	-0.169** (0.062)	-0.192*** (0.046)	-0.191*** (0.046)
Main genre of developer	0.156** (0.059)	0.168** (0.059)	0.121* (0.060)	0.133* (0.060)	0.019 (0.050)	0.025 (0.050)
Sales of publisher (ln)	0.113*** (0.023)	0.110*** (0.022)	0.005 (0.021)	0.001 (0.020)	-0.069*** (0.018)	-0.070*** (0.018)
Platform-PlayStation	0.048 (0.076)	0.052 (0.076)	0.072 (0.096)	0.077 (0.095)	0.041 (0.083)	0.043 (0.083)
Platform-Xbox	0.065 (0.078)	0.073 (0.078)	-0.114 (0.102)	0.105 (0.101)	-0.156† (0.091)	-0.153† (0.091)
Platform-Nintendo	-0.120 (0.111)	-0.109 (0.109)	0.087 (0.106)	0.099 (0.104)	0.165† (0.091)	0.169† (0.091)
Platform-PC	0.472*** (0.091)	0.472*** (0.091)	-0.153 (0.127)	-0.152 (0.127)	-0.460*** (0.109)	-0.458*** (0.109)
Critics evaluations					0.652*** (0.021)	0.648*** (0.021)
Constant	8.226*** (0.491)	8.505*** (0.505)	8.507*** (0.341)	8.808*** (0.330)	3.146*** (0.442)	3.295*** (0.430)
Year of release FE	Yes	Yes	Yes	Yes	Yes	Yes
Genre FE	Yes	Yes	Yes	Yes	Yes	Yes
ESRB rating FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	4645	4645	4645	4645	4645	4645

FE: fixed-effects; ESRB: Entertainment Software Rating Board.

Standard errors in parentheses clustered by game developers.

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$; † $p < 0.10$.

market audiences (Santos et al., 2019). Table 2 presents the results of Tobit regressions. Model 1 and Model 3 show the baseline results including only control variables. As expected, *developer reputation* and *developer status* are positively associated with the evaluations of both critics and consumers. *Genre fuzziness* is negatively related to critic evaluations, suggesting that developing games in fuzzy categories leads to less favorable evaluations from critics. The effect of *genre fuzziness* is, however, non-significant on consumer evaluations, which is somehow aligned with prior proposition that categorical ambiguity may have different impacts on different types of audiences (Pontikes, 2012).

Models 2 and 4 include *signal inconsistency* of a developer's past game reputation. The effect of signal inconsistency is significantly negative on both critic evaluations ($\beta = -1.651$, *s.e.* = 0.425) and consumer evaluations ($\beta = -1.780$, *s.e.* = 0.510). This supports the baseline *Hypothesis 1* that signal inconsistency leads to negative evaluations among both critic and consumer audiences. To formally compare the two coefficients across the two models, I use seemingly unrelated estimations (*suest* in Stata), which simultaneously estimate (co)variance (Ertug et al., 2016). I then employ *ncom* (Haans et al., 2016) to calculate the ratio between the effect size of *signal inconsistency* in Model 4 and that in Model 2. A ratio that is larger than 1 would mean that the former's effect size is stronger than the latter's, and vice versa.¹³ However, I find that the ratio is not significantly different from 1 ($p = 0.750$), suggesting that signal inconsistency has comparable *total* effects on the evaluations of critics and customers.

However, since signal inconsistency may influence consumers through the mediating effect of critics, it is necessary to control for critic evaluations, to test the *direct* effect of signal inconsistency on consumers. As such, I include critic evaluations in Models 5 and 6.¹⁴ The effect of critic evaluations on consumer evaluations is positive in Model 5 ($\beta = 0.652$, *s.e.* = 0.021), suggesting that consumers are influenced by critics. The effect of signal inconsistency on consumer evaluations decreases substantially in Model 6 ($\beta = -0.710$) from Model 4 ($\beta = -1.780$). The ratio between the effect sizes of *signal inconsistency* in Model 6 and in Model 2 is 0.43 ($= -0.710 / -1.651$). It means that the direct effect size of signal inconsistency on consumer evaluations is less than half of its direct effect on critic evaluations. The formal *ncom* test also suggests that the ratio is significantly smaller than one ($p = 0.021$). This lends support to *Hypothesis 2* that the negative effect of signal inconsistency on audience evaluations is weaker for consumers than for critics.

To test Hypothesis 3 on the mediation effect of critic evaluations, I employ SEM estimation. Table 3 presents the SEM results. The overall findings are consistent with my Tobit regressions in Table 2. Model 1 shows the total effects on consumer evaluations. Model 2 and Model 3 show the direct effect of signal inconsistency on consumer evaluations and critic evaluations, respectively, and Model 4 reports the indirect effect on consumer evaluations. The direct effect of signal inconsistency on consumer evaluations ($\beta = -0.703$, *s.e.* = 0.391) counts for only about 39.6% of its total effect ($\beta = -1.773$, *s.e.* = 0.509), while the indirect effect ($\beta = -1.070$, *s.e.* = 0.378) through critic evaluations takes a much larger proportion (60.4%). To more formally test the mediation effect, I used Preacher and Hayes' (2004) calculation tool for various forms of Sobel tests. The classic Sobel test is highly significant ($t = -3.85$; $p < 0.001$), while the Aroian and Goodman specifications provide almost identical results. These results together suggest that the effect of signal inconsistency on consumer evaluations is largely mediated by critic evaluations. *Hypothesis 3* is hence supported that critic evaluations mediate the effect of signal inconsistency on consumer evaluations.¹⁵

Supplementary analyses

In addition to main estimation, I conduct a wide range of supplementary analyses that examine alternative measurements, use alternative model specification, and further explore the data.

Standardization. Instead of using raw measures of dependent variables, I standardize them before regressions. Much of my analytical attention is focused on comparing estimated coefficients. While the independent variable (i.e. *signal inconsistency*) stays the same across models, dependent variables are different. One may be concerned that my results were driven by the fact that

Table 3. SEM of signal inconsistency on critic and consumer evaluations.

Variables	Total effect on consumer evaluations	Direct effect on consumer evaluations	Direct effect on critic evaluations	Indirect effect on consumer evaluations
	Model 1	Model 2	Model 3	Model 4
Signal inconsistency	-1.773*** (0.509)	-0.703 [†] (0.391)	-1.649*** (0.425)	-1.070*** (0.278)
Critic evaluations	0.649*** (0.021)	0.649*** (0.021)		
Developer reputation	0.380*** (0.073)	-0.142* (0.067)	0.804*** (0.082)	0.522*** (0.059)
Developer status	0.280** (0.105)	-0.137 (0.088)	0.643*** (0.100)	0.417*** (0.069)
No. of genres (ln)	0.211** (0.074)	0.034 (0.057)	0.273** (0.086)	0.177** (0.056)
Niche density (ln)	0.023 (0.045)	0.009 (0.032)	0.022 (0.039)	0.014 (0.025)
Genre fuzziness	-0.610 (0.710)	0.432 (0.612)	-1.607* (0.727)	-1.042* (0.475)
Multi-homing	-0.168** (0.061)	-0.191*** (0.046)	0.035 (0.066)	0.023 (0.042)
Sales of publisher (ln)	0.000 (0.021)	-0.071*** (0.017)	0.111*** (0.022)	0.072*** (0.015)
Platform-PlayStation	0.078 (0.093)	0.047 (0.082)	0.048 (0.075)	0.031 (0.049)
Platform-Xbox	-0.105 (0.100)	-0.151 (0.091)	0.071 (0.078)	0.046 (0.051)
Platform-Nintendo	0.099 (0.102)	0.172 (0.090)	-0.113 (0.109)	-0.073 (0.071)
Platform-PC	-0.151 (0.127)	-0.457*** (0.109)	0.472*** (0.091)	0.306*** (0.061)
Main genre of developer	0.111** (0.038)		0.172** (0.059)	0.111** (0.038)
Year of release FE	Yes	Yes	Yes	Yes
Genre FE	Yes	Yes	Yes	Yes
ESRB rating FE	Yes	Yes	Yes	Yes
Observations	4645	4645	4645	4645

FE: fixed-effects; ESRB: Entertainment Software Rating Board.
Standard errors in parentheses clustered by game developers.
*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$; [†] $p < 0.10$.

dependent variables have different distributions and scales. To address this concern, I standardize both critic and consumer evaluations before estimation, and report the SEM results in Table 4. The regression results and Sobel tests ($t = -3.85$) are both largely consistent with Table 3, in term of statistical significance and effect size.

Signal inconsistency. I use an alternative measure of signal inconsistency. In my main estimation, the inconsistency of reputation signals is measured based on critic ratings. This measure is valid as

Table 4. Supplementary analyses—standardized DVs..

Variables	Total effect on consumer evaluations	Direct effect on consumer evaluations	Direct effect on critic evaluations	Indirect effect on consumer evaluations
	Model 1	Model 2	Model 3	Model 4
Signal inconsistency	-1.255*** (0.360)	-0.498† (0.277)	-1.239*** (0.319)	-0.757*** (0.197)
Critic evaluations	0.611*** (0.020)	0.611*** (0.020)		
Developer reputation	0.269*** (0.052)	-0.101* (0.047)	0.604*** (0.062)	0.369*** (0.042)
Developer status	0.198** (0.074)	-0.097 (0.062)	0.483*** (0.075)	0.295*** (0.049)
No. of genres (ln)	0.149** (0.052)	0.024 (0.041)	0.205** (0.065)	0.125** (0.039)
Niche density (ln)	0.016 (0.032)	0.006 (0.023)	0.016 (0.029)	0.010 (0.018)
Genre fuzziness	-0.432 (0.503)	0.306 (0.433)	-1.207* (0.546)	-0.738* (0.337)
Multi-homing	-0.119** (0.043)	-0.135*** (0.032)	0.027 (0.049)	0.016 (0.030)
Sales of publisher (ln)	0.000 (0.015)	-0.051*** (0.012)	0.083*** (0.017)	0.051*** (0.010)
Platform-PlayStation	0.055 (0.066)	0.033 (0.058)	0.036 (0.057)	0.022 (0.035)
Platform-Xbox	-0.074 (0.071)	-0.107† (0.064)	0.054 (0.059)	0.033 (0.036)
Platform-Nintendo	0.070 (0.072)	0.122† (0.064)	-0.085 (0.082)	-0.052 (0.050)
Platform-PC	-0.107 (0.090)	-0.324*** (0.077)	0.355*** (0.069)	0.217*** (0.043)
Main genre of developer	0.079** (0.027)		0.129** (0.044)	0.306*** (0.061)
Year of release FE	Yes	Yes	Yes	Yes
Genre FE	Yes	Yes	Yes	Yes
ESRB rating FE	Yes	Yes	Yes	Yes
Observations	4645	4645	4645	4645

DV: dependent variable; FE: fixed-effects; ESRB: Entertainment Software Rating Board.

Standard errors in parentheses clustered by game developers.

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$; † $p < 0.10$.

prior research commonly refers to critic ratings as an important quality reputation indicator of video games (Rietveld and Eggers, 2018). However, one may be concerned that the critics-based inconsistency measure mechanically leads to a stronger impact on critic evaluations than on consumer evaluations. To address this concern, I replace the measure of signal inconsistency using the standard deviation of *user ratings* of games in the past 5 years, scaled by its mean ratings. This new measure has a moderate correlation with the critic-based measure ($r=0.373$).¹⁶ The regression results are reported in Table 5, which are largely similar to my main findings above.

Table 5. Measuring inconsistency with consumer ratings.

Variables	Total effect on consumer evaluations	Direct effect on consumer evaluations	Direct effect on critic evaluations	Indirect effect on consumer evaluations
	Model 1	Model 2	Model 3	Model 4
Signal inconsistency	-0.931** (0.316)	-0.326 (0.227)	-0.943*** (0.243)	-0.606*** (0.158)
Critic evaluations	0.642*** (0.021)	0.642*** (0.021)		
Developer reputation	0.443*** (0.068)	-0.101 (0.065)	0.847*** (0.078)	0.544*** (0.056)
Developer status	0.245* (0.105)	-0.150 (0.088)	0.614*** (0.104)	0.395*** (0.071)
No. of genres (ln)	0.171* (0.073)	0.025 (0.057)	0.226** (0.084)	0.145** (0.054)
Niche density (ln)	0.040 (0.046)	0.019 (0.033)	0.033 (0.040)	0.021 (0.025)
Genre fuzziness	-0.578 (0.736)	0.469 (0.644)	-1.630* (0.738)	-1.047* (0.477)
Multi-homing	-0.161** (0.061)	-0.191*** (0.046)	0.047 (0.065)	0.030 (0.042)
Sales of publisher (ln)	-0.005 (0.021)	-0.075*** (0.018)	0.109*** (0.023)	0.070*** (0.015)
Platform-PlayStation	0.046 (0.094)	0.033 (0.085)	0.019 (0.077)	0.012 (0.049)
Platform-Xbox	-0.143 (0.101)	-0.167 (0.093)	0.036 (0.079)	0.023 (0.051)
Platform-Nintendo	0.069 (0.102)	0.162 (0.092)	-0.144 (0.112)	-0.093 (0.072)
Platform-PC	-0.161 (0.126)	-0.454*** (0.110)	0.457*** (0.094)	0.293*** (0.062)
Main genre of developer	0.102** (0.038)		0.159** (0.059)	0.102** (0.038)
Year of release FE	Yes	Yes	Yes	Yes
Genre FE	Yes	Yes	Yes	Yes
ESRB rating FE	Yes	Yes	Yes	Yes
Observations	4520	4520	4520	4520

FE: fixed-effects; ESRB: Entertainment Software Rating Board.

Standard errors in parentheses clustered by game developers.

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$.

Moreover, I also build a variable that captures the “second order” discrepancy between critic- and consumer-based inconsistency measures (formally, $|$ critic-based inconsistency/consumer-based inconsistency $-1|$). A larger value of discrepancy indicates that the inconsistency measures are more divergent between critic and consumer sides. The results in Table 6 show that the effect of the discrepancy is non-significant in all models. This suggests that market audiences are not very attentive to this “second-order” factor.

Table 6. Discrepancy between critic and consumer-side inconsistency.

Variables	Total effect on consumer evaluations	Direct effect on consumer evaluations	Direct effect on critic evaluations	Indirect effect on consumer evaluations
	Model 1	Model 2	Model 3	Model 4
Signal inconsistency	0.019 (0.015)	0.008 (0.015)	0.016 (0.020)	0.011 (0.013)
Critic evaluations	0.645*** (0.021)	0.645*** (0.021)		
Developer reputation	0.486*** (0.072)	-0.085 (0.066)	0.884*** (0.083)	0.571*** (0.060)
Developer status	0.256* (0.106)	-0.149 (0.088)	0.627*** (0.105)	0.405*** (0.072)
No. of genres (ln)	0.142 (0.075)	0.014 (0.056)	0.199* (0.089)	0.128* (0.057)
Niche density (ln)	0.041 (0.046)	0.020 (0.033)	0.033 (0.040)	0.021 (0.026)
Genre fuzziness	-0.819 (0.741)	0.434 (0.640)	-1.941* (0.773)	-1.253* (0.504)
Multi-homing	-0.172** (0.062)	-0.197*** (0.046)	0.039 (0.067)	0.025 (0.043)
Sales of publisher (ln)	0.001 (0.022)	-0.073*** (0.018)	0.115*** (0.023)	0.074*** (0.015)
Platform-PlayStation	0.045 (0.097)	0.038 (0.086)	0.010 (0.077)	0.007 (0.050)
Platform-Xbox	-0.146 (0.104)	-0.164 (0.094)	0.028 (0.080)	0.018 (0.051)
Platform-Nintendo	0.065 (0.106)	0.165 (0.093)	-0.155 (0.114)	-0.100 (0.074)
Platform-PC	-0.165 (0.129)	-0.453*** (0.111)	0.447*** (0.093)	0.288*** (0.062)
Main genre of developer	0.102** (0.039)		0.158** (0.060)	0.102** (0.039)
Year of release FE	Yes	Yes	Yes	Yes
Genre FE	Yes	Yes	Yes	Yes
ESRB rating FE	Yes	Yes	Yes	Yes
Observations	4497	4497	4497	4497

FE: fixed-effects; ESRB: Entertainment Software Rating Board.
Standard errors in parentheses clustered by game developers.

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$.

Market sales. I use global sales of video games as an alternative measure of consumer evaluations. Sales reflect how video games are received by end consumers in the marketplace (Zhao et al., 2018). Using game sales helps differentiate the measurement methods of two dependent variables, critic and consumer evaluations. This new measure also provides insights into the selection process. In other words, while online ratings reflect consumers' post-consumption evaluations, sales indicate the patterns in pre-consumption evaluations. I report the results in Table 7, which are largely consistent with my findings in Table 3. Specifically, the total effect of signal consistency on game sales is largely driven by the indirect effect through critic evaluations (with a significant Sobel test of $t = -3.70$).

Table 7. Measuring consumer evaluation with global sales.

Variables	Total effect on consumer evaluations	Direct effect on consumer evaluations	Direct effect on critic evaluations	Indirect effect on consumer evaluations
	Model 1	Model 2	Model 3	Model 4
Signal inconsistency	-0.266* (0.126)	-0.062 (0.114)	-1.649*** (0.425)	-0.204*** (0.055)
Critic evaluations	0.124*** (0.010)	0.124*** (0.010)		
Developer reputation	0.186*** (0.043)	0.086* (0.042)	0.804*** (0.082)	0.100*** (0.012)
Developer status	0.325* (0.131)	0.246* (0.124)	0.643*** (0.100)	0.080*** (0.014)
No. of genres (ln)	0.065* (0.032)	0.031 (0.031)	0.273** (0.086)	0.034** (0.011)
Niche density (ln)	0.061*** (0.014)	0.058*** (0.012)	0.022 (0.039)	0.003 (0.005)
Genre fuzziness	-0.002 (0.285)	0.197 (0.282)	-1.607* (0.727)	-0.199* (0.090)
Multi-homing	-0.043 (0.030)	-0.047 (0.029)	0.035 (0.066)	0.004 (0.008)
Sales of publisher (ln)	0.078*** (0.008)	0.065*** (0.007)	0.111*** (0.022)	0.014*** (0.003)
Platform-PlayStation	0.204*** (0.041)	0.198*** (0.039)	0.048 (0.075)	0.006 (0.009)
Platform-Xbox	0.137*** (0.036)	0.128*** (0.034)	0.071 (0.078)	0.009 (0.010)
Platform-Nintendo	0.167* (0.078)	0.181* (0.074)	-0.113 (0.109)	-0.014 (0.014)
Platform-PC	-0.106* (0.048)	-0.164*** (0.046)	0.472*** (0.091)	0.058*** (0.012)
Main genre of developer	0.021** (0.008)		0.172** (0.059)	0.021** (0.008)
Year of release FE	Yes	Yes	Yes	Yes
Genre FE	Yes	Yes	Yes	Yes
ESRB rating FE	Yes	Yes	Yes	Yes
Observations	4645	4645	4645	4645

FE: fixed-effects; ESRB: Entertainment Software Rating Board.

Standard errors in parentheses clustered by game developers.

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$.

Developer-level evidence. I also extend my analysis to the developer level. As my primary research design is at the game-platform level, there are multiple games from the same developers at the same year, who share the same signal inconsistency. This may lead to concerns about the interdependence of observations. In this extension, I aggregate game ratings into the developer level, which helps reduce interdependence of observations. Specifically, I calculate the average ratings of a developer's all games in the same year and analyze how a developer's signal inconsistency affects its overall evaluations in the market. The results are reported in Table 8, which show patterns consistent with my core claims.

Table 8. Supplementary analysis—developer-level evidence.

Variables	Ave. of critic evaluations		Ave. of consumer evaluations	
	Model 1	Model 2	Model 2	Model 3
Signal inconsistency	-1.276** (0.460)	-1.703** (0.575)		-0.894* (0.444)
Developer reputation	0.948*** (0.077)	0.349*** (0.100)		-0.252* (0.098)
Developer status	0.720*** (0.125)	0.397* (0.169)		-0.060 (0.132)
Fuzziness	-2.526*** (0.694)	-0.847 (0.868)		0.756 (0.750)
Ave. of niche density	-0.146** (0.048)	0.040 (0.059)		0.133** (0.051)
Sales of publisher (ln)	0.119*** (0.022)	-0.035 (0.025)		-0.111*** (0.023)
Ave. of critic evaluations				0.634*** (0.032)
Constant	7.673*** (0.409)	8.408*** (0.331)		3.541*** (0.436)
Year of release FE	Yes	Yes		Yes
Observations	2555	2555		2555

FE: fixed-effects.

Standard errors in parentheses clustered by game developers.

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$.

Game quality. I try to rule out the impact of unobserved quality. It is possible that key variables are all driven by game quality that is not captured in my estimation. In this extension, I follow prior studies (Rietveld and Eggers, 2018) to focus on a subsample of multi-homing games. The same game released on multi-platforms entails almost identical quality. However, its developer could have different signal inconsistency in different platforms, when game ratings differ across platforms even for the same game titles (Zhu and Zhang, 2010). To do so, I first create developer-platform level measure of signal inconsistency. That is, while my primary inconsistency measure is at the developer level, this new measure calculates the standard deviation of game ratings for each developer on each platform in the past 5 years. Despite its advantage of controlling for game quality, it is worth noting that this design has a strong assumption that there is no information spillover across platforms. Reputation signals on one platform are not perceived by audiences in the other platforms.

After adopting this approach, I end up with a small subsample of multi-homing games with sufficient data (929 observations of 415 games). I then apply generalized SEM with game-level fixed-effects to estimate critic and consumer evaluations, which rules out game-level and developer-level factors. The results are reported in Table 9. The overall picture is consistent with my hypotheses, with t -statistic of -1.84 from Sobel test. However, it is worth noting that the effects are less substantive in this estimation, as the key variables are mostly significant at the $p < 0.1$ level. These preliminary results may imply that the effect of signal inconsistency on audience evaluations is also largely explained by game quality. New games from a signal-inconsistent developer tend to receive negative evaluations, not only because signal inconsistency creates ambiguity in the

Table 9. Supplementary analysis—game quality.

Variables	Total effect on consumer evaluations	Direct effect on consumer evaluations	Direct effect on critic evaluations	Indirect effect on consumer evaluations
	Model 1	Model 2	Model 3	Model 4
Signal inconsistency	−0.930 [†] (0.522)	−0.620 (0.498)	−0.667 [†] (0.356)	−0.310 [†] (0.169)
Critic evaluations (std)	0.465*** (0.045)	0.465*** (0.045)		
Niche density (ln)	0.260** (0.083)	0.176* (0.076)	0.180** (0.055)	0.084** (0.028)
Xbox	0.002 (0.144)	−0.151 (0.133)	0.329*** (0.098)	0.153** (0.049)
PS	−0.014 (0.144)	−0.132 (0.133)	0.254** (0.096)	0.118* (0.047)
Constant	6.990*** (0.739)	2.981*** (0.745)	8.621*** (0.382)	4.009** (0.410)
Year FE	Yes	Yes	Yes	Yes
Observations	929	929	929	929
Number of groups	415	415	415	415

FE: fixed-effects.

Standard errors in parentheses clustered by games.

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$; [†] $p < 0.10$.

market (a perception discount), but also because those games are likely to have lower inherent quality (a production discount). In other words, in bridging the effect of signal inconsistency on consumers, critics may have both integrated their evaluative concerns about the inconsistency and helped reveal the hidden product quality associated with signal inconsistency. Nonetheless, to understand the actual relationship between quality and signal inconsistency, more direct theorizing and testing are needed.

Specialists versus generalists. To further understand the role of signal inconsistency, I conduct a set of subsample analyses that distinguish between specialist and generalist developers. My theory suggests that consumer audiences are less likely to attend to all signals of a developer. It would be particularly so if the developer is a generalist who diversifies into many game genres, assuming that end consumers tend to focus on a narrow set of game genres. In other words, it is even more unlikely for consumers to sense a generalist developer's various signals in many different genres, as compared with a specialist's signals. To test this corollary, I divide the full sample into two based on the median of the number of genres that a developer spans in the past 5 years, and then repeat SEM estimations in each of the two subsamples. In Table 10, analyzing specialist developers, the direct effect of signal inconsistency on consumer evaluations is negative and significant ($\beta = -1.286$, $s.e. = 0.481$), whose effect size is comparable to that on critic evaluations. In Table 11, analyzing generalist developers, however, the direct effect on consumer evaluations is non-significant ($\beta = 0.220$, $s.e. = 0.577$), whereas there is a significant direct effect on critic evaluations ($\beta = -2.514$, $s.e. = 0.667$). While inconclusive, comparing the two results suggests that the effect of signal inconsistency on consumer audiences is weaker when game developers are more of generalists diversifying into many genres.

Table 10. Supplementary analysis—specialist developers.

Variables	Total effect on consumer evaluations	Direct effect on consumer evaluations	Direct effect on critic evaluations	Indirect effect on consumer evaluations
	Model 1	Model 2	Model 3	Model 4
Signal inconsistency	-2.087** (0.683)	-1.286** (0.481)	-1.204* (0.555)	-0.801* (0.373)
Critic evaluations	0.665*** (0.028)	0.665*** (0.028)		
Developer reputation	0.441*** (0.104)	-0.145 (0.097)	0.881*** (0.090)	0.586*** (0.067)
Developer status	0.190 (0.286)	-0.332 (0.218)	0.784*** (0.186)	0.521*** (0.123)
No. of genres (ln)	0.372 (0.223)	0.190 (0.184)	0.274 (0.200)	0.182 (0.133)
Niche density (ln)	0.008 (0.055)	0.003 (0.044)	0.008 (0.045)	0.005 (0.030)
Genre fuzziness	-0.789 (0.952)	0.326 (0.843)	-1.676* (0.806)	-1.115* (0.542)
Multi-homing	-0.194* (0.080)	-0.206** (0.068)	0.018 (0.073)	0.012 (0.048)
Sales of publisher (ln)	-0.030 (0.024)	-0.077*** (0.023)	0.071** (0.022)	0.047** (0.015)
Platform-PlayStation	0.139 (0.140)	0.100 (0.110)	0.059 (0.110)	0.039 (0.074)
Platform-Xbox	-0.012 (0.144)	-0.075 (0.114)	0.094 (0.111)	0.062 (0.074)
Platform-Nintendo	0.227 (0.158)	0.299* (0.120)	-0.109 (0.146)	-0.072 (0.097)
Platform-PC	0.029 (0.171)	-0.302* (0.146)	0.498*** (0.125)	0.332*** (0.085)
Main genre of developer	0.061 (0.053)		0.092 (0.081)	0.061 (0.053)
Year of release FE	Yes	Yes	Yes	Yes
Genre FE	Yes	Yes	Yes	Yes
ESRB rating FE	Yes	Yes	Yes	Yes
Observations	2325	2325	2325	2325

FE: fixed-effects; ESRB: Entertainment Software Rating Board.
Standard errors in parentheses clustered by game developers.

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$.

Consumer experience. The final extension focuses on intra-audience heterogeneity. My main framework emphasizes inter-audience distinction. Specifically, I posit that critics are more sensitive to signal inconsistency because they hold a more holistic view on diverse signals than consumers. If so, one may expect differences among consumers across levels of experience: more experienced consumers may react more strongly to signal inconsistency than less experienced consumers. To test this argument, I collect consumer-game level ratings. Consumer experience is simply measured as the number of game ratings posted by a consumer before the focal game. I then

Table 11. Supplementary analysis—generalist developers.

Variables	Total effect on consumer evaluations	Direct effect on consumer evaluations	Direct effect on critic evaluations	Indirect effect on consumer evaluations
	Model 1	Model 2	Model 3	Model 4
Signal inconsistency	-1.376* (0.665)	0.220 (0.577)	-2.514*** (0.667)	-1.596*** (0.423)
Critic evaluations	0.635*** (0.029)	0.635*** (0.029)		
Developer reputation	0.208 (0.123)	-0.144 (0.112)	0.553*** (0.140)	0.351*** (0.092)
Developer status	0.306*** (0.077)	-0.025 (0.070)	0.520*** (0.101)	0.330*** (0.069)
No. of genres (ln)	0.287* (0.121)	0.002 (0.077)	0.448** (0.147)	0.284** (0.094)
Niche density (ln)	0.026 (0.068)	0.004 (0.048)	0.034 (0.061)	0.021 (0.039)
Genre fuzziness	-0.530 (1.093)	0.559 (0.850)	-1.715 (1.329)	-1.089 (0.843)
Multi-homing	-0.141 (0.092)	-0.160* (0.063)	0.029 (0.109)	0.018 (0.069)
Sales of publisher (ln)	0.055 (0.033)	-0.061** (0.022)	0.183*** (0.040)	0.116*** (0.026)
Platform-PlayStation	0.040 (0.124)	0.027 (0.113)	0.021 (0.101)	0.013 (0.064)
Platform-Xbox	-0.175 (0.133)	-0.192 (0.129)	0.027 (0.107)	0.017 (0.068)
Platform-Nintendo	0.023 (0.122)	0.101 (0.113)	-0.122 (0.135)	-0.077 (0.086)
Platform-PC	-0.362 (0.196)	-0.633*** (0.161)	0.427** (0.146)	0.271** (0.097)
Main genre of developer	0.177** (0.054)		0.278*** (0.082)	0.177** (0.054)
Year of release FE	Yes	Yes	Yes	Yes
Genre FE	Yes	Yes	Yes	Yes
ESRB rating FE	Yes	Yes	Yes	Yes
Observations	2320	2320	2320	2320

FE: fixed-effects; ESRB: Entertainment Software Rating Board.

Standard errors in parentheses clustered by game developers.

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$.

run *consumer-game* level estimations to examine the role of consumer experience. Results are presented in Table 12. Models 1 and 2 employ Tobit regressions, where I find a significant and negative interaction between consumer experience and signal inconsistency on individual ratings ($\beta = -0.583$, s.e. = 0.087). It suggests that consumers with more rating experiences are indeed more reactive to the inconsistency of reputation signals. Models 3 and 4 use game-level fixed-effects that rule out game- and developer-level factors.¹⁷ The interaction effect in Model 4 is significant and negative on individual ratings. Altogether, the primary results show patterns that are in line with my conjecture.

Table 12. Supplementary analysis—consumer-game level estimations.

Variables	Tobit		Game FE	
	Model 1	Model 2	Model 3	Model 4
Signal inconsistency \times consumer experience		-0.583*** (0.087)		-0.223** (0.078)
Consumer experience	-0.006 (0.006)	0.061*** (0.012)	-0.058*** (0.006)	-0.032** (0.011)
Signal inconsistency	0.398** (0.146)	1.040*** (0.182)		
Critics evaluations	0.599*** (0.009)	0.599*** (0.009)		
Developer reputation	-0.237*** (0.026)	-0.234*** (0.026)		
Developer status	-0.031 (0.030)	-0.027 (0.030)		
No. of genres (ln)	-0.147*** (0.022)	-0.144*** (0.022)		
Niche density (ln)	0.049** (0.016)	0.049** (0.016)		
Genre fuzziness	-0.450 [†] (0.238)	-0.457 [†] (0.238)		
Multi-homing	-0.311*** (0.022)	-0.311*** (0.022)		
Main genre of developer	-0.098*** (0.020)	-0.098*** (0.020)		
Sales of publisher (ln)	-0.136*** (0.007)	-0.137*** (0.007)		
Platform-PlayStation	-0.086 [†] (0.045)	-0.085 [†] (0.045)		
Platform-Xbox	-0.285*** (0.046)	-0.283*** (0.046)		
Platform-Nintendo	0.288*** (0.049)	0.290*** (0.049)		
Platform-PC	-1.145*** (0.053)	-1.143*** (0.053)		
Constant	-0.408* (0.161)	-0.469** (0.161)	7.352*** (0.011)	7.352*** (0.011)
Year of release FE	Yes	Yes	No	No
Genre FE	Yes	Yes	No	No
ESRB rating FE	Yes	Yes	No	No
Observations	110,372	110,372	110,609	110,609
R-squared			0.268	0.268

FE: fixed-effects; ESRB: Entertainment Software Rating Board.

Heteroskedasticity-consistent standard errors in parentheses.

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$; [†] $p < 0.10$.

Discussion

Moving beyond the baseline proposition that signal inconsistency leads to evaluation penalty (Vergne et al., 2018; Zhao and Zhou, 2011), I draw attention to a key scope condition for signal inconsistency to take effect. Specifically, I argue that the effect of signal inconsistency is highly

contingent on the extent to which audiences attend to and interpret various market signals. Signal inconsistency is hence proposed to exert a stronger direct effect on critics than consumers, because consumer audiences are much less likely to keep a full track of various signals than critics. I test these hypotheses in a sample of video games. The results show that while the direct effects of signal inconsistency on critic evaluations and consumer evaluations are both negative, the former's effect size is more than two times as large as the latter's. I also find a substantial mediating effect of critic evaluations on the relationship between signal inconsistency and consumer evaluations.

This study contributes mainly to the signaling literature. While the role of market signals has been well established, recent literature shifts attention to the important issue of signal inconsistency, with a holistic perspective on multiple signals from the same entity (Stern et al., 2014). Signal inconsistency leads to penalty because audiences find it challenging to make sense of incongruent quality signals. While scholars have underscored the negative consequences of signal inconsistency (Parker et al., 2017; Wang and Jensen, 2019), most studies in this stream focus on producers and the diverse signals they send, at the cost of treating audiences as homogeneous. This study adds to this literature on signal inconsistency by emphasizing the important role of audiences who process different signals. Specifically, I stress that signals are not equally observed by any audiences. Audiences are not only divergent in their tastes and preference (Kim and Jensen, 2011; Shrum, 1991), but also different in the extent to which they attend to and interpret market signals (Kovács and Hannan, 2010). When audiences fail to consider different signals from the same entity, they may not sense its signal inconsistency.

Supporting my arguments, I find that signal inconsistency has a stronger effect on critic evaluations than consumer evaluations. This implies the importance of considering heterogeneity at the audience side (i.e. signal receivers) when theorizing about signal inconsistency. Such findings are also in line with recent proposition that critics attend to a different set of information cues than consumers (Falchetti et al., 2022). However, it does not mean that critic evaluations are more accurate than consumer evaluations, since new products with inconsistent past signals may not necessarily have worse inherent quality. To understand evaluation accuracy, one would have to measure inherent quality, which can be challenging in the cultural market.

I also highlight the market mediation role of professional critics in processing signal inconsistency. Interestingly, while the direct effect of signal inconsistency is much smaller on consumer evaluations than on critic evaluations, its total effects are comparable on both. It suggests that although consumers may not clearly observe signal inconsistency themselves, they are influenced by signal inconsistency indirectly when they accept the evaluation schema of critics (Hsu et al., 2012b). This provides additional evidence on the important role of critics as intermediaries that reduce information asymmetry in the market. Critics search for, select, and process various vital signals in their evaluations, which facilitates the quality judgments of consumers who otherwise would overlook such important cues as signal inconsistency.

By doing so, this study underscores the importance of controlling for mutual influence between different audiences. While research has proliferated on how different types of audiences react differently to organization practices and market signals (Cattani et al., 2014), many studies in this stream do not highlight the interactions between different audiences. Such neglect is a bit surprising, since it has also been well established that different groups of audiences often affect each other's evaluations (Kovács and Sharkey, 2014; Zhao et al., 2018). In this study, I find that after controlling for the influence of critics on consumers, the effect size of signal inconsistency on consumer evaluation is reduced by about 60%. That is, the influence of signal inconsistency on consumer evaluations is largely explained by the mediating effect of critic evaluations. This suggests the importance of accounting for the mutual influence between different groups of audiences, when analyzing how they respond to organizational practices and market signals. For instance,

even though investors and customers prefer different types of startups (Pontikes, 2012), investors' decisions may also be influenced by whether startups will be well accepted by customers. When investors foresee that an atypical startup will be less favorably received by customers, will they still invest in the startup? Interdependence between audiences is simple in this study, as it is more likely for consumers to be influenced by critics, than the other way around. It can become, however, complicated for other types of audiences where the influence is largely bi-directional.

Limitations and future studies

This research opens more space for future studies. First, although I motivate this project with market signal inconsistency in general, I develop and test hypotheses on a very specific form of signal inconsistency, that is, the inconsistency of quality signals over time. Future studies may extend it in two different ways. On one hand, scholars may examine other forms of quality signal inconsistency, such as inconsistency across different dimensions (Zhao and Zhou, 2011), inconsistency across different product categories (Wang and Jensen, 2019), and even inconsistency between different groups of audience evaluation. On the other hand, while most of current attention is paid to the inconsistency of vertical orderings, it is also important to explore the inconsistency of horizontal/categorical signals (Negro et al., 2015). For instance, when a firm makes distant jumps in its categorical focus over time, it may face the problem of categorical identity inconsistency (Leung, 2014). Although those are different forms of inconsistency, I suspect the underlying theory may still hold that the effect of signal inconsistency depends on the extent to which audiences attend to various market signals. Nonetheless, more empirical investigation is needed.

Second, while this study focuses on the distinction between critics and consumers (i.e. inter-audience heterogeneity), I have downplayed the heterogeneity within each type of audiences (i.e. intra-audience heterogeneity). Critics (or consumers) in fact also differ in terms of their attention, experience, taste, or purpose (Goldberg et al., 2016), which may lead to divergent evaluations among themselves. In other words, whereas my research analyzes aggregate evaluations of critics (or consumers), it is interesting to examine the variance of their individual evaluations (Sun, 2012). Although I touch upon it in additional analyses, more work is needed. Zooming into individual evaluations also allows scholars to explore their dynamics and inconsistency at the product (game) level. Some games encounter a high consensus among audiences, while others a high dissensus (d'Astous and Touil, 1999). It is therefore interesting to investigate the implications of this type of inconsistency at the product level, which is overlooked in this firm-level inconsistency research. Relatedly, while focusing on the aggregate evaluations of audiences, I have also missed their dynamics. Earlier reviews may shape how later audiences evaluate focal products (Hoskins et al., 2021; Pang et al., 2022). I leave this question to future studies.

Third, whereas I emphasize that audiences differ in their attention to market signals, I do not look into their heterogeneity in taste and preference (Holbrook, 1999; Kim and Jensen, 2011). Future studies could try to explore whether certain types of audiences are in favor of signal inconsistency. Such research will help us understand the potential positive side of signal inconsistency, extending the current focus on its negative consequences. Relatedly, while I compare critics and consumers, two important audiences in the market, it would be interesting to extend my framework to other types of audiences, such as investors, peers, and strategic partners (Cattani et al., 2014; Pontikes, 2012). For instance, how would investors or competitors react to the inconsistency of a firm's reputation signals?

Finally, it is important to note that my findings can be context-dependent. In evaluating new games, for instance, consumers usually follow critics (Zhu and Zhang, 2010), which is an important scope condition for my mediation hypothesis. In the settings where consumers and critics

evaluations are simultaneous, however, the argument for the mediation role of critics might not hold. Moreover, in the setting of video games, consumers pay close attention to critic evaluations, so that their behavior is influenced by the latter (Zhao et al., 2018). If consumers in other settings rely more on their own assessments or other social media, rather than referring to critics as their opinion leaders (Eliashberg and Shugan, 1997), the mediating role of critics will be dampened. It is also useful to consider the distinction between connoisseurial and procedural reviews (Blank, 2007). My empirical evidence is based on the connoisseurial reviews of video games that are subject to audiences' experiences, tastes, and training. In the process of procedural reviews (e.g. digital cameras or software) that depends on objective and mechanical testing, we may see a different pattern.

Conclusion

In conclusion, this study emphasizes a key scope condition for signal inconsistency to take effects: audiences process various signals emitted by a firm. Based on that, I argue and find evidence that professional critics are more reactive to signal inconsistency than end consumers in their evaluations of video games. I also observe that critics act as an important market intermediary to bridge the effect of signal inconsistency on consumers. Moving beyond the general signal inconsistency discount, this article highlights the importance of investigating audience heterogeneity in the research of signal inconsistency.

Acknowledgements

The author would like to thank the Coeditor Oliver Alexy and three anonymous reviews for their constructive feedback and guidance in the review process, which substantially improved the manuscript.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iD

Pengfei Wang  <https://orcid.org/0000-0002-5594-5787>

Notes

1. While research on cognition heuristics suggests that people may not have to possess perfect knowledge for signal inconsistency to remain important (Goldstein and Gigerenzer, 2002), it is still plausible that audiences with a more complete coverage of signals are more likely to accurately perceive the level of inconsistency (Connelly et al., 2011; Madsen and Rodgers, 2015).
2. It is useful to elaborate on the antecedents of signal inconsistency. First, inconsistency may result from uncertainties in new product development, such that developers are unable to introduce products with consistent or stable quality. Second, inconsistency may be related to developers' different strategies. When a developer aims to hunt for blockbusters, it is more likely to experience inconsistency, as compared with developers going for strong and stable average quality. Moreover, inconsistency in a large market may be a simple reflection of regression to the mean. In this study I focus on the consequences of signal inconsistency, and hence downplay the heterogeneity in its antecedents. However, exploring the antecedents may have important implications, deserving more dedicated further studies.
3. This baseline hypothesis assumes keeping consistent the average quality of developers. While the hypothesis is intuitive for developers with high- or middle-level average quality, it may be less so for low-quality developers. In other words, the inconsistency for a low-reputation developer might mean that it can sometimes introduce good games. It is less problematic for my analysis, as developers in the

sample are mostly of high- or middle-level quality (see more in the empirics). However, this also means that my findings may be not generalizable to very low-quality developers. It is therefore necessary for further studies to examine how the relationship changes over different average quality levels, using more comprehensive data.

4. Audiences' theory of value is defined as how audiences identify issues and solutions, ascribe value, and rank solution providers (Durand and Boulongne, 2017). In my context, consumers' theory of value therefore refers to how end consumers ascribe value to games.
5. A relevant question here is: if consumers pay close attention to critic evaluations, why should they evaluate video games themselves? There are at least two main theories to explain it. First, psychological motivation theory suggests that consumers share their evaluations online for self-enhancement. Second, review environment theory posits that consumers' decisions to post evaluations are subject to environmental factors such as others' opinions. See more in Zhou and Duan (2016).
6. It is worth noting that critics may mediate the effect of signal inconsistency on consumer evaluations, in both pre- and post-consumption periods. In the pre-consumption stage when consumers select games to purchase, they are likely to be influenced by critics, largely because of quality uncertainty associated with experience goods (Zhao et al., 2018). In the post-consumption stage when consumers need to rate the games they have purchased, they face less uncertainty about the features of the games. However, consumers may still be unsure about what ratings the games deserve. As such, they may still (explicitly or implicitly) engage in critics' judgments and recommendations when making their post-consumption evaluations. Indeed, Hoskins et al. (2021) show that in the US beer product category, consumer evaluations are likely to follow professional critics even after they have tasted the beers.
7. I limit my focus on a sample of video games with global sales greater than 10,000 copies because smaller games receive much less attention from the market and are less relevant in directing the evaluations of critics and consumers. The sample covers about 30% of games in the database for the period. The number of games analyzed per year in my sample is comparable to prior studies on video games (Rietveld and Eggers, 2018).
8. Such sequence features are more observed in the project-based settings such as video games or films (Eliashberg and Shugan, 1997; Zhu and Zhang, 2010). In other settings where temporal sequence is more ambiguous, it would be more difficult to examine the mediation effect of critics on consumers. For instance, in evaluating restaurants that are not project-based, critics and consumers may post (or update) their reviews anytime, with no clear sequence of their reaction timelines.
9. I choose the 20th percentile, according to the statistics in Han and Pollock (2021: 98). In unreported analysis, I also use alternative cutoffs (e.g. 10th, 25th percentile) and find similar results. Moreover, the results also stay similar if excluding this variable.
10. I also restrict my sample to the three dominant platforms: Xbox, PlayStation, and Nintendo. The results are stable.
11. ESRB is a self-regulated organization that categorizes games based on their contents. Games are reviewed and assigned different ratings such as EC (early childhood), T (teen), and AO (adults only).
12. Because my data are structured at the game-platform level (i.e. there is only one observation per game-platform), rather than as time series at the developer level, serial autocorrelation is not a major concern. In other words, whereas the dependent variable is the evaluation of a new game, the independent variable is the dispersion of past games' ratings. That is, the two variables are measured for different game subjects, instead of the same games. However, those observations are not completely independent as some games are created by the same developer. Therefore, I cluster standard errors at the game developer level.
13. More formally, I use the following equation: $\ln \text{com_b[consumer evaluations]signal inconsistency} / _b[\text{critic evaluations}] \text{signal inconsistency} - 1$.
14. In these models, the direct effect of main genre on user ratings is non-significant, suggesting that it is a valid instrument (at least empirically) for my SEM.
15. In the SEM results, the direct effect of signal inconsistency on critic ratings is around 2.3 times ($= -1.649 / -0.703$) as large as its direct effect on user ratings, which is similar to the results in my separate regressions in Table 2.

16. Due to data availability of consumer ratings, the sample size (4520) is slightly smaller than my primary sample (4645).
17. However, because signal inconsistency is also constant at the game level, its direct effect cannot be separately estimated in Model 3. Please also note that the consumer-game level analysis is still preliminary because it does not consider how a consumer's rating is influenced by the critic and consumer ratings prior to that.

References

- Askin N and Bothner MS (2016) Status-aspirational pricing: The “Chivas Regal” strategy in US higher education, 2006–2012. *Administrative Science Quarterly* 61(2): 217–253.
- Basuroy S, Abraham Ravid S, Gretz RT, et al. (2020) Is everybody an expert? An investigation into the impact of professional versus user reviews on movie revenues. *Journal of Cultural Economics* 44(1): 57–96.
- Becker HS (2008) *Art Worlds: Updated and Expanded*. Berkeley, CA: University of California Press.
- Blank G (2007) *Critics, Ratings, and Society: The Sociology of Reviews*. Lanham, MD: Rowman & Littlefield.
- Cattani G, Ferriani S and Allison PD (2014) Insiders, outsiders, and the struggle for consecration in cultural fields: A core-periphery perspective. *American Sociological Review* 79(2): 258–281.
- Chen Y, Liu Y and Zhang J (2012) When do third-party product reviews affect firm value and what can firms do? The case of media critics and professional movie reviews. *Journal of Marketing* 76(2): 116–134.
- Choi HS, Ko MS, Medlin D, et al. (2018) The effect of intrinsic and extrinsic quality cues of digital video games on sales: An empirical investigation. *Decision Support Systems* 106: 86–96.
- Connelly BL, Certo ST, Ireland RD, et al. (2011) Signaling theory: A review and assessment. *Journal of Management* 37(1): 39–67.
- d'Astous A and Touil N (1999) Consumer evaluations of movies on the basis of critics' judgments. *Psychology & Marketing* 16(8): 677–694.
- Davila A, Foster G and Gupta M (2003) Venture capital financing and the growth of startup firms. *Journal of Business Venturing* 18(6): 689–708.
- De Vaan M, Stark D and Vedres B (2015) Game changer: The topology of creativity. *American Journal of Sociology* 120(4): 1144–1194.
- Durand R and Boulongne R (2017) Advancing category research: Theoretical mapping and under-researched areas. In: Greenwood R, Oliver C, Sahlin K, et al. (eds) *The Sage Handbook of Organizational Institutionalism*. Thousand Oaks, CA: SAGE, pp. 647–668.
- Durand R and Haans RF (2022) Optimally distinct? Understanding the motivation and ability of organizations to pursue optimal distinctiveness (or not). *Organization Theory*. Epub ahead of print 17 March. DOI: 10.1177/26317877221079341.
- Ebbers JJ and Wijnberg NM (2010) Disentangling the effects of reputation and network position on the evolution of alliance networks. *Strategic Organization* 8(3): 255–275.
- Eliashberg J and Shugan SM (1997) Film critics: Influencers or predictors? *Journal of Marketing* 61(2): 68–78.
- Ertug G and Castellucci F (2013) Getting what you need: How reputation and status affect team performance, hiring, and salaries in the NBA. *Academy of Management Journal* 56(2): 407–431.
- Ertug G, Yogev T, Lee YG, et al. (2016) The art of representation: How audience-specific reputations affect success in the contemporary art field. *Academy of Management Journal* 59(1): 113–134.
- Falchetti D, Cattani G and Ferriani S (2022) Start with “why,” but only if you have to: The strategic framing of novel ideas across different audiences. *Strategic Management Journal* 43(1): 130–159.
- Goldberg A, Hannan MT and Kovács B (2016) What does it mean to span cultural boundaries? Variety and atypicality in cultural consumption. *American Sociological Review* 81(2): 215–241.
- Goldstein DG and Gigerenzer G (2002) Models of ecological rationality: The recognition heuristic. *Psychological Review* 109(1): 75.
- Haans RF, Pieters C and He ZL (2016) Thinking about U: Theorizing and testing U- and inverted U-shaped relationships in strategy research. *Strategic Management Journal* 37(7): 1177–1195.
- Han J-H and Pollock TG (2021) The two towers (or somewhere in between): The behavioral consequences of positional inconsistency across status hierarchies. *Academy of Management Journal* 64(1): 86–113.

- Holbrook MB (1999) Popular appeal versus expert judgments of motion pictures. *Journal of Consumer Research* 26(2): 144–155.
- Hoskins J, Gopinath S, Verhaal JC, et al. (2021) The influence of the online community, professional critics, and location similarity on review ratings for niche and mainstream brands. *Journal of the Academy of Marketing Science* 49(6): 1065–1087.
- Hsu G (2006a) Evaluative schemas and the attention of critics in the US film industry. *Industrial and Corporate Change* 15(3): 467–496.
- Hsu G (2006b) Jacks of all trades and masters of none: Audiences' reactions to spanning genres in feature film production. *Administrative Science Quarterly* 51(3): 420–450.
- Hsu G, Hannan MT and Koçak Ö (2009) Multiple category memberships in markets: An integrative theory and two empirical tests. *American Sociological Review* 74(1): 150–169.
- Hsu G, Negro G and Perretti F (2012a) Hybrids in Hollywood: A study of the production and performance of genre-spanning films. *Industrial and Corporate Change* 21(6): 1427–1450.
- Hsu G, Roberts PW and Swaminathan A (2012b) Evaluative schemas and the mediating role of critics. *Organization Science* 23(1): 83–97.
- Jensen M and Roy A (2008) Staging exchange partner choices: When do status and reputation matter? *Academy of Management Journal* 51(3): 495–516.
- Jensen M and Wang P (2018) Not in the same boat: How status inconsistency affects research performance in business schools. *Academy of Management Journal* 61(3): 1021–1049.
- Jensen M, Kim H and Kim BK (2012) Meeting expectations: A role-theoretic perspective on reputation. In: Barnett ML and Pollock TG (eds) *The Oxford Handbook of Corporate Reputation*. Oxford: Oxford University Press, pp. 140–159.
- Kim BK and Jensen M (2011) How product order affects market identity: Repertoire ordering in the US opera market. *Administrative Science Quarterly* 56(2): 238–256.
- Kim H and Jensen M (2014) Audience heterogeneity and the effectiveness of market signals: How to overcome liabilities of foreignness in film exports? *Academy of Management Journal* 57(5): 1360–1384.
- Kovács B and Hannan MT (2010) The consequences of category spanning depend on contrast. In: Hsu G, Negro G and Koçak Ö (eds) *Categories in Markets: Origins and Evolution*. Bingley: Emerald Group Publishing Limited.
- Kovács B and Sharkey AJ (2014) The paradox of publicity: How awards can negatively affect the evaluation of quality. *Administrative Science Quarterly* 59(1): 1–33.
- Leack J (2015) Opencritic's gamer-centric style is everything metacritic should have been. Available at: <http://www.gamerevolution.com/features/11820-opencritics-gamercentric-style-is-everything-metacritic-should-have-been/> (accessed 4 October 2022).
- Leung MD (2014) Dilettante or renaissance person? How the order of job experiences affects hiring in an external labor market. *American Sociological Review* 79(1): 136–158.
- Madsen PM and Rodgers ZJ (2015) Looking good by doing good: The antecedents and consequences of stakeholder attention to corporate disaster relief. *Strategic Management Journal* 36(5): 776–794.
- Milgrom P and Roberts J (1986) Price and advertising signals of product quality. *Journal of Political Economy* 94(4): 796–821.
- Negro G, Hannan MT and Fassiotto M (2015) Category signaling and reputation. *Organization Science* 26(2): 584–600.
- Nelson P (1970) Information and consumer behavior. *Journal of Political Economy* 78(2): 311–329.
- Ni SB and Tang Y (2013) Which signal to rely on? The impact of the quality of board interlocks and inventive capabilities on research and development alliance formation under uncertainty. *Strategic Organization* 11(4): 364–388.
- Obloj T and Capron L (2011) Role of resource gap and value appropriation: Effect of reputation gap on price premium in online auctions. *Strategic Management Journal* 32(4): 447–456.
- Pang J, Liu AX and Golder PN (2022) Critics' conformity to consumers in movie evaluation. *Journal of the Academy of Marketing Science* 50: 864–887.
- Parker ON, Krause R and Covin JG (2017) Ready, set, slow: How aspiration-relative product quality impacts the rate of new product introduction. *Journal of Management* 43(7): 2333–2356.

- Podolny JM (1993) A status-based model of market competition. *American Journal of Sociology* 98(4): 829–872.
- Pollock TG and Gulati R (2007) Standing out from the crowd: The visibility-enhancing effects of IPO-related signals on alliance formation by entrepreneurial firms. *Strategic Organization* 5(4): 339–372.
- Pontikes EG (2012) Two sides of the same coin: How ambiguous classification affects multiple audiences' evaluations. *Administrative Science Quarterly* 57(1): 81–118.
- Preacher KJ and Hayes AF (2004) SPSS and SAS procedures for estimating indirect effects in simple mediation models. *Behavior Research Methods, Instruments, & Computers* 36: 717–731.
- Reinstein DA and Snyder CM (2005) The influence of expert reviews on consumer demand for experience goods: A case study of movie critics. *The Journal of Industrial Economics* 53(1): 27–51.
- Remo C (2008) Microsoft to delist low-ranking XBLA titles, raise size limit. Available at: <https://www.gamedeveloper.com/pc/microsoft-to-delist-low-ranking-xbla-titles-raise-size-limit> (accessed 4 October 2022).
- Rhee M and Haunschild PR (2006) The liability of good reputation: A study of product recalls in the US automobile industry. *Organization Science* 17(1): 101–117.
- Rietveld J (2018) Creating and capturing value from freemium business models: A demand-side perspective. *Strategic Entrepreneurship Journal* 12(2): 171–193.
- Rietveld J and Eggers J (2018) Demand heterogeneity in platform markets: Implications for complementors. *Organization Science* 29(2): 304–322.
- Santos T, Lemmerich F, Strohmaier M, et al. (2019) What's in a review: Discrepancies between expert and amateur reviews of video games on metacritic. *Proceedings of the ACM on Human-Computer Interaction* 3(CSCW): 1–22.
- Shapiro C (1983) Premiums for high quality products as returns to reputations. *The Quarterly Journal of Economics* 98(4): 659–679.
- Sharkey A, Kovacs B and Hsu G (2022) Expert critics, rankings, and review aggregators: The changing nature of intermediation and the rise of markets with multiple intermediaries. *Academy of Management Annals*. Epub ahead of print 15 April. DOI: 10.5465/annals.2021.0025.
- Shaver JM (2005) Testing for mediating variables in management research: Concerns, implications, and alternative strategies. *Journal of Management* 31(3): 330–353.
- Shrum W (1991) Critics and publics: Cultural mediation in highbrow and popular performing arts. *American Journal of Sociology* 97(2): 347–375.
- Singh J and Fleming L (2010) Lone inventors as sources of breakthroughs: Myth or reality? *Management Science* 56(1): 41–56.
- Sorenson O (2014) Status and reputation: Synonyms or separate concepts? *Strategic Organization* 12(1): 62–69.
- Spence M (1974) *Market Signaling: Informational Transfer in Hiring and Related Screening Processes*. Cambridge: Harvard University Press.
- Steigenberger N and Wilhelm H (2018) Extending signaling theory to rhetorical signals: Evidence from crowdfunding. *Organization Science* 29(3): 529–546.
- Stern I, Dukerich JM and Zajac E (2014) Unmixed signals: How reputation and status affect alliance formation. *Strategic Management Journal* 35(4): 512–531.
- Stuart K (2015) Has video game reviewing become an impossible task? Available at: <https://www.theguardian.com/technology/2015/nov/25/video-game-reviewing-critics-industry> (accessed 4 October 2022).
- Sun M (2012) How does the variance of product ratings matter? *Management Science* 58(4): 696–707.
- Vergne JP, Wernicke G and Brenner S (2018) Signal incongruence and its consequences: A study of media disapproval and CEO overcompensation. *Organization Science* 29(5): 796–817.
- Wang P and Jensen M (2019) A bridge too far: Divestiture as a strategic reaction to status inconsistency. *Management Science* 65(2): 859–878.
- Zhao EY, Ishihara M, Jennings PD, et al. (2018) Optimal distinctiveness in the console video game industry: An exemplar-based model of proto-category evolution. *Organization Science* 29(4): 588–611.
- Zhao W and Zhou X (2011) Status inconsistency and product valuation in the California wine market. *Organization Science* 22(6): 1435–1448.

- Zhou W and Duan W (2016) Do professional reviews affect online user choices through user reviews? An empirical study. *Journal of Management Information Systems* 33(1): 202–228.
- Zhu F and Zhang X (2010) Impact of online consumer reviews on sales: The moderating role of product and consumer characteristics. *Journal of Marketing* 74(2): 133–148.
- Zuckerman EW (1999) The categorical imperative: Securities analysts and the illegitimacy discount. *American Journal of Sociology* 104(5): 1398–1438.
- Zuckerman EW and Kim TY (2003) The critical trade-off: Identity assignment and box-office success in the feature film industry. *Industrial and Corporate Change* 12(1): 27–67.

Author biography

Pengfei Wang is an Associate Professor in BI Norwegian Business School. He obtained his PhD in Strategy from Erasmus University Rotterdam. His research covers social status, market categories, and technological innovation. His studies have been accepted for publication by journals including *Academy of Management Journal*, *Management Science*, *Research Policy*, *Journal of Business Venturing*, and *Industrial and Corporate Change*.

Appendix I

Example game developers with their games and ratings.

Farsight (High inconsistency)			Game developer: Level 5 (Low inconsistency)		
Game titles	Platform	Ratings	Game titles	Platform	Ratings
Hard Rock Casino	PSP	44	Professor Layton and the Curious Village	DS	85
Pinball Hall of Fame: The Williams Collection	PSP	78	Professor Layton and the Diabolical Box	DS	84
Pinball Hall of Fame: The Williams Collection	Wii	82	White Knight Chronicles: International Edition	PS3	64
Game Party 3	Wii	37	Professor Layton and the Unwound Future	DS	86
Pinball Hall of Fame: The Williams Collection	PS3	82	Dragon Quest IX: Sentinels of the Starry Skies	DS	87
Pinball Hall of Fame: The Williams Collection	X360	83	Professor Layton and the Last Specter	DS	83
Brunswick Pro Bowling	PS3	47	White Knight Chronicles II	PS3	60
Brunswick Pro Bowling	X360	44	Ni no Kuni: Wrath of the White Witch	PS3	85
Game Party in Motion	X360	33	Inazuma Eleven Strikers	Wii	59

This table shows the example games from two game developers, Farsight Studios from the United States and Level 5 from Japan. Apparently, neither of them has been offering games with constant quality ratings. However, Farsight seems to be more inconsistent than Level 5 in terms of ratings of their games.

Appendix 2

Sequence of critic and user reviews

Take the game *Inside* on the Xbox platform (released on 29 June 2016). As of May 2022, there are in total 87 critic ratings on Metacritic. Out of 87 ratings, 41 are published before its official release date. Another 25 ratings are published within 1 week after the release. To gain more understandings about review timelines, I collect information on the dates of critic reviews for a sample of games released between 2011 and 2016. I find that critics swiftly rate games upon (if not before) their releases, with a median of 7 days after release dates. By contrast, the majority of user ratings are published over 1 week after the releases. Indeed, according to the extensive analysis of Metacritic reviews by Santos et al. (2019), users post game ratings in a median of 65 days after releases. As such, there exists a disparity between the evaluation timelines of critics and consumers.

Analyzing further the pattern of reviews in the sample, I also find that the overall critic ratings used in my main analysis are highly correlated with critic ratings before game releases ($r=0.90$), and highly correlated with critic ratings within 1 week upon game releases ($r=0.94$). This suggests that the overall critic evaluation on a game has been largely reflected by their early ratings, even though some critics publish ratings at a slightly late period. Certainly, it is possible that a few user ratings come before some late critic ratings, such that the latter can be influenced by the former (Pang et al., 2022). But those cases are less frequent or substantive, so that they would not shift the overall evaluation pattern too much. In other words, although there might be a small proportion of overlaps between critic and user ratings, it is still plausible to assume that critic evaluations are largely presented prior to user evaluations. Indeed, because of the temporal sequence, consumer behavior has been constantly found to be shaped by critic evaluations in this setting (Zhao et al., 2018; Zhu and Zhang, 2010), rather than the other way around.