Does Synchronicity of Emotion Between Steamers and Viewers Influence Consumption? Evidence from Live Streaming of Virtual Idols

Menghan Duan
Fudan University
21110690007@m.fudan.edu.cn

Yueyue Zhang
University of Nottingham Ningbo China
yueyue.zhang@nottingham.edu.cn

Qi Zhang
Southwest Jiaotong University
qi.zhang@swjtu.edu.cn

Cheng Zhang
Fudan University
zhangche@fudan.edu.cn

Abstract

Live streaming has become increasingly popular for broadcasting live events or experiences to an online audience in real time. With the prevalence of the metaverse, virtual idols appear on the stage of live streaming. Unlike real human streamers, virtual idols' appearance is computer-generated and does not exist in the physical world. Though existing literature has documented the effect of emotions on viewers' behavior, they mainly focus on real humans and either side of streamers or viewers. To fill this gap, we consider both sides and focus on the synchronicity of emotion between the two sides, which we term steamer-viewer emotional synchronicity. We compare the relationship between steamer-viewer emotional synchronicity and viewers' consumption across the live stream by virtual idols vs. real humans. Our analysis of large-scale video data shows that compared to real-human streamers, virtual streamers are more likely to trigger steamer-viewer emotional synchronicity when expressing happiness, leading to more consumption; but virtual streamers struggle more to do so when conveying sadness, leading to reduced consumption. This study sheds light on how avatars can be designed to leverage emotion to engage viewers.

Keywords: Virtual idol, Emotional synchronicity, Live streaming, Video mining, Metaverse.

1. Introduction

The concept of the metaverse has been around for decades. Recent advancements in virtual and augmented reality technologies have made the metaverse a more tangible possibility. After Facebook changed its name to Meta, a series of well-known companies have also subsequently entered this realm. Metaverse depicts a virtual world or universe that is created by the convergence of physical and digital spaces, wherein users are expected to interact with the projected identities of other users (avatars) and objects (Lee and Chen 2011). The use of avatars for representing and acting on behalf of people as their digital twins within metaverse environments is an integral part of the metaverse (Dwivedi et al. 2022).

There is growing interest in avatar-based business practices, including the development of virtual idols or virtual influencers. Generated by computers, though, one specific type of virtual idol is controlled by anonymous human performers wearing motion-capture gear. These virtual idols show their talents like singing and dancing in front of millions of viewers on live-streaming platforms. In response, fans might tip money and make purchases. For example, A-SOUL, a virtual idol group that debuted in December 2020, became China's most influential virtual idol in just one year. The superior motion-capture technology, visual design, and 5G network were considered the key reasons why A-SOUL has outweighed other virtual idols. Because we focus on virtual idol streamers that are performed by real humans, the major difference between these virtual idols and real human idols lies in their appearance. Virtual idols have entirely digital appearances and are not real people. They can have fantastical or unique appearances that may not be achievable in the real world. Real human idols are real people with their own unique appearances. In addition, the voice of virtual idols is from real humans

URI: https://hdl.handle.net/10125/106730
978-0-9981331-7-1

performing behind an animated face rather than generated by software[1].

Research firm iiMedia says the virtual idol industry in China earned 3.5 billion yuan ($547 million) in 2020, with year-on-year growth of 70.3%[2]. Virtual idol live-streamers have outperformed real human hosts by 3 times in terms of engagement on China's video platform Bilibili as of 2020[1]. As of 2022, there were 230,000 active virtual streamers, and the number of broadcasts reached 29.2 billion[3]. Given the popularity of virtual idols in business, viewers' interactions with virtual idols and the monetization of virtual idols, however, receive relatively scarce scholarly attention.

Live streaming effectively bridges streamers (i.e., content creators) with their audience, collaboratively fostering a unique online community. Within this community, emotional exchanges play a pivotal role in communication. A burgeoning stream of work has delved into the role of emotions in live-streaming sales (e.g., Bharadwaj et al. 2022; Lin et al. 2021; Teixeira et al. 2012; Wang et al. 2022). However, these studies mainly focus on real human streamers and the emotion of either the streamer-side or the viewer-side of the live stream. In this study, we consider both sides and focus on the synchronicity of emotion between the two sides, which we term steamer-viewer emotional synchronicity. It refers to the phenomenon where individuals experience similar or synchronized emotions during an interaction or over time. Meanwhile, we investigate a novel and increasingly important context in which the streamers are virtual idols. We contend that virtual idols and real humans may trigger different levels of steamer-viewer emotional synchronicity, thus leading to different in-consumption outcomes.

Specifically, we address the following knowledge gaps in this study: 1). Does steamer-viewer emotional synchronicity exist in live streaming? 2). If yes, does steamer-viewer emotional synchronicity vary across live streams by real-human vs. virtual idol? and 3). Does steamer-viewer emotional synchronicity impact viewers' consumption behavior? To investigate these issues, we focus on the live streaming of virtual idols and conduct deep learning approaches to investigate the interactions between streamers and their viewers based on moment-to-moment data.

2. Literature and hypotheses development

In this section, we review relevant works and highlight the potential contributions of this study. Relevant works include virtual idols, live streaming, and the roles of emotions in live streaming.

2.1. Virtual idols in live streaming

Live streaming typically refers to the real-time transmission of video and audio content over the Internet (Bharadwaj et al. 2022). Viewers could interact with streamers by commenting and virtual gifting (Lin et al. 2021; Wohn and Freeman 2020). To understand the effectiveness of live streaming, scholars have investigated the effects of emotional displays of streamers. For example, scholars have started to apply artificial intelligence technologies to automatically detect a streamer's face in each video frame, extract emotional expressions, and relate them to sales performance (e.g., Bharadwaj et al. 2022). However, extant studies have mostly focused on live streaming by real humans and in the retail context.

We contend that live streaming by virtual idols differs from that by real humans in two major aspects. First, human faces could be easily distinguished from avatars. In the virtual idol context, viewers see computer-generated characters or animated faces that do not exist in the physical world. Second, the consumption content and interface presented are varied. In retailing, streamers use screen-mediated sales presentations for a variety of products and marketing communications. Viewers are persuaded to purchase products as introduced by the streamers (or sellers). By contrast, virtual idols who are designed to be entertainers often appear in video games, anime, and other forms of digital media. Viewers can consume such entertainment content for free. Viewers' intention to purchase or pay only emerges when they appreciate the performance and when they are eager to engage in more immersive interaction through commenting and virtual gifting (Wohn and Freeman 2020). Different from acquiring virtual goods in online games for players' own sake, buying virtual goods or tipping in live streaming is for streamers and is a way of social interaction and exchange (Xu et al. 2022).

Due to these differences in human appearance and live-streaming content, we are wondering whether and how the virtual idol streamers' emotional expressions affect viewers' behavior, especially viewers' tipping behavior. Existing literature on live streaming may not provide sufficient guidance to answer this question. In a first effort to fill this void, we empirically examine the

[1] Source: https://www.scmp.com/lifestyle/entertainment/article/3111890/chinas-virtual-idols-digital-avatars-and-human-voices-meet

[2] Source: https://news.cgtn.com/news/2021-10-29/China-s-virtual-idols-worth-big-money-in-metaverse-era-14Kr8PeLRja/index.html

[3] Source: https://finance.sina.cn/2023-01-19/detail-imyasyin5546896.d.html

effects of emotion in the live streaming of virtual idols on viewers' consumption.

2.2. Moment-to-moment (MTM) synchronicity in emotion

Because viewers can react to video content at any moment, it enables us to capture viewers' consumption, engagement, and interactions with virtual idol streamers by analyzing moment-to-moment data (Zhang et al. 2020). Live streaming is featured as a dyadic process of communication. Consumers' (viewers') engaging activities take place throughout a live stream, and multiple payments might occur (Lin et al. 2021). Scholars have long analyzed the role of emotion in buyer-seller exchange relationships. On the buyer side, Teixeira et al. (2012) offer the first study to automatically extract the emotions of joy and surprise that viewers experience when watching television commercials and then relate these emotions to attention and ad avoidance behavior. Liu et al. (2018) incorporate prospective consumers' emotional displays into their study. They examined the impact of emotional displays of viewers on sales using the Facial Action Coding System and categorized a set of emotions (e.g., happiness, surprise, disgust) based on the facial expressions of viewers watching movie trailers. They find that viewers' emotional displays of happiness positively influence both watching intentions and box-office revenues. On the seller side of the exchange dyad, Bharadwaj et al. (2022) are the first to assess the sales impact of product, price, sales force, advertising, and promotion in the presence of a salesperson's face and emotional displays. They find that salespeople's emotions negatively impact sales across all six emotions, including happiness.

Most of the previous research on live streaming focuses on either side of viewers or streamers (e.g., Bharadwaj et al. 2022; Teixeira et al. 2012). However, Limited research has tapped into the moment-to-moment (MTM) synchrony between streamers and viewers to combine supply-side dynamics with demand-side data.

Synchrony depicts the underlying temporal structure of highly aroused moments of interpersonal exchange. A relevant study has examined the synchronicity between the volume of live comments and movie content variations at the MTM level (Zhang et al., 2020). In their study, though, synchronicity is based on content co-movement, while the synchronicity of emotion co-movement between the two sides is not clarified. There is another study that considers the interactive and dynamic settings of live streaming and examines how viewers' underlying emotion is likely to be affected by the broadcaster's emotion (Lin et al. 2021). Though this study has investigated interdependencies and coevolution of viewer and broadcaster behaviors in live streaming, the subject is mainly broadcasters with real-human appearances. However, virtual idols with animated faces differ from real humans in appearance, which may lead to a different observation of the effect of emotional synchronicity in a live-streaming setting.

2.3. Streamer-viewer emotional synchronicity in live streaming

Emotion is an immediate and transient object-focused motivational response broadly linked to the valence and arousal dimensions and is expressed in a bivariate manner (Barrett et al. 2007; Eckler and Bolls 2011; Lang and Bradley 2010). The affective dimension of customer experience pertains to the elicitation of moods and emotions as a result of interactions with a firm (Brakus et al., 2009). Emotional synchronicity often occurs in social interactions, and highlights the interconnectedness of emotions in social settings. It refers to the phenomenon where people display emotions similar to that observed in others with whom they have had contact (e.g., Byrne and Russon 1998; Chartrand and Bargh 1999). It means that emotions can align or resonate between people, leading to a shared emotional experience. There is increasing evidence that such social influence can occur, at least in part, in the absence of such conscious deliberations (e.g., Aarts et al. 2008). In our setting, the perception of streamers' emotion can trigger overlapping representations of that emotion in the viewers, leading to synchronicity without intention or awareness.

Emotional synchronicity can involve both positive and negative emotions. For example, people may experience shared joy or shared sadness during a specific moment. Positive emotions are often linked to favorable feedback and proactive interactions, thus eliciting more immediate responses. When a streamer radiates emotions of happiness, it aligns well with the foundational essence upon which the community was built. As active participants in this community, the audience often resonates with these positive emotions, leading to feelings of joy, excitement, and satisfaction, as they fully engage in this uplifting atmosphere.

When a streamer expresses feelings of sadness, it goes beyond superficial emotions and is perceived by the audience as a genuine experience that the streamer is sharing with them. As a result, the audience can connect with the streamer on a deeper level, acknowledging their humanity. This often leads to a heightened sense of emotional connection between the streamer and viewers, fostering greater emotional synchronicity.

Given the real-time nature of live streaming, the emotion expressed by the streamers may be instantaneously echoed within the community every moment. Based on the aforementioned arguments, we contend that both levels of happiness and sadness expressed by streamers are related to the streamer-viewer emotional synchronicity in the community. Therefore, we hypothesize that:

Hypothesis 1a: The streamer's emotion of happiness positively influences the streamer-viewer emotional synchronicity.

Hypothesis 1b: The streamer's emotions of sadness positively influence the streamer-viewer emotional synchronicity.

2.4. Emotional synchronicity and viewers' consumption

Emotional synchronicity can influence viewers' consumption for two reasons. First, emotional synchronicity amplifies social connections and a sense of belonging. According to research by Batson et al. (1991), emotional resonance can enhance concern for others, making individuals more inclined to harbor desires to assist others. This kind of synchronicity lays the groundwork for a trusting relationship, making viewers more likely to form an emotional bond with the streamer.

Second, rooted in the Social Exchange Theory (Lawler and Thye 1999), individuals often seek equilibrium in social interactions. When they derive some form of value from others, such as emotional support or an entertainment experience, they might be compelled to reciprocate to maintain this balance. The perceived emotional value from streamers might affect viewers' motivation in relationships, leading to an urge to reciprocate streamers through more content consumption. Therefore, we hypothesize that:

Hypothesis 2a: The streamer-viewer emotional synchronicity in happiness positively influences tipping from viewers.

Hypothesis 2b: The streamer-viewer emotional synchronicity in sadness positively influences tipping from viewers.

2.5. The moderating effect of human vs. virtual appearance

When evaluating the differences between virtual streamers and real-human streamers, a crucial consideration is how they are perceived by viewers. In the context of live streaming, social presence becomes a more pronounced factor that differentiates virtual idols and real human streamers. Researchers have discovered that social presence can exert an influence on consumers' purchase intentions by providing a sense of human warmth and sociability (Ye et al., 2020). Numerous empirical studies have demonstrated that the absence of social interactions impedes consumers' online purchase intentions (e.g., Lu et al., 2016). This is because social cues, which are capable of generating a feeling of social presence, significantly impact consumers in their decision-making process regarding purchases (Khwaja et al., 2019).

Studies have found that the positive emotions of live streamers attract viewers to reward them. Situational characteristics of human vs. virtual streamers might moderate the relationship. Direct interaction with a human streamer may increase the salience of the emotional experience as a result of rapport between streamer and audience, and thus increases the influence of the respective emotion on evaluation or paying behavior.

In sad scenarios that are grounded in trust, virtual streamers might find it more challenging to secure deep trust from viewers compared to their real human counterparts. The emotions exhibited by real-human streamers are rooted in their genuine experiences and sentiments, a reality that viewers might find easier to accept and resonate with. In contrast, though virtual streamers can realistically mimic the outward display of emotions, they may be perceived by viewers as superficial, and lacking in genuine life experiences due to the low level of social presence. Such superficiality might reduce the trust viewers have in virtual streamers, reducing the affective arousal and willingness to consume.

Hypothesis 3a: The relationship between a streamer's happiness and streamer-viewer emotional synchronicity is stronger for real human streamers than for virtual idol streamers.

Hypothesis 3b: The relationship between a streamer's sadness and streamer-viewer emotional synchronicity is stronger for real human streamers than for virtual idol streamers.

Figure 1 below illustrates the research model.

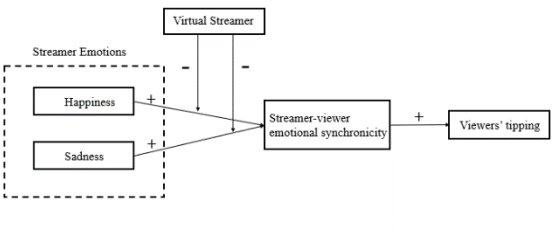

Figure 1. Research Design

3. Data and methodology

3.1. Recognition of streamers' and viewers' emotions

To investigate the effect of emotional synchronicity, we identified streamers' and viewers' emotions.

3.1.1. Acoustic emotions in streamers' speech. To accurately recognize emotional nuances in Chinese speech, we employed a fine-tuned version of the Wav2Vec 2.0 base model (Baevski et al. 2020). The wav2vec is an algorithm oriented towards speech classification tasks, demonstrating good performance in both accuracy and efficiency (Baevski et al. 2020). The dataset used for fine-tuning hailed from the CAISA Chinese Speech Emotion dataset (http://www.chineseldc.org/resource_info.php?rid=76), comprising 1,200 recordings performed by actors in Chinese, embodying six distinct emotional states: happiness, sadness, surprise, anger, fear, and neutrality. The concept of six primary or universal emotions, which often include happiness, sadness, surprise, anger, fear, and neutrality (or sometimes referred to as disgust or contempt), is a simplified framework used to describe and categorize human emotional experiences. According to Paul Ekman's basic emotions theory, the six primary emotions are universal across cultures and are associated with distinct facial expressions.

3.1.2. Textual emotions in viewers' danmaku. The text content of interest in our study is the viewers' danmaku. The term "danmaku" (also named "bullet screen comment" or "post" in English) refers to a unique and interactive form of user-generated commentary that appears as scrolling text comments or messages overlaid on the video being streamed. To identify emotions from texts, we first used "jieba" (Sun 2012), an open-source Chinese word segmentation toolkit, to segment the sentences. A list of stop words is used to remove words that are unlikely to assist in topic modeling like common words, modal particles, prepositions, connectives, and so on. It is noted that this step is performed separately for the streamers and the viewers. Then, we performed the multi-classification prediction to distinguish the six emotions: happiness, sadness, surprise, anger, fear, and neutrality. Considering the text in live streaming is short, we selected the pretrained dataset, SMP-2020-EWECT (https://smp2020ewect.github.io/), for fine-tuning our text-emotion classifier. We tried several state-of-the art models such as Bert (Devlin et al. 2018) and Ernie (Zhang et al. 2019) with varied model parameters to compare their accuracy. Table 1 reports the performance comparison. It turns out that the paddle-Ernie-3.0 model achieves the best performance. Finally, we applied the model to obtain viewers' danmaku emotions for our sample.

Table 1. Algorithm Comparison for Emotion Classification in Text

Model	Bert	Bert-bilstm	Erinetiny	Erinetiny -2.0	Erine3.0
Accuracy	43%	52%	65%	80%	84%

3.2. Emotional synchronicity measurement

Leveraging both the streamer voice emotion recognition algorithm and the barrage (danmaku) text emotion recognition algorithm, we obtained a six-dimensional emotion vector for the barrage at each interval (with each interval lasting 20 seconds). Given that there's a time lag in the audience's reaction to the streamer's emotion, we offset the barrage text emotion vector by one interval. Consequently, we employed the cosine similarity between the streamer's voice emotion vector at interval t and the viewers' danmaku text emotion vector at interval t+1 to gauge the degree of emotional synchronicity between the streamer and the viewers at the interval.

3.3. Viewers' tipping measurement

We obtained the tipping records of viewers through the live streaming platform's API. Specifically, there are two main forms of tipping by the audience: sending paid danmaku and purchasing virtual gifts. We measure the amount of tipping during the t-period of live stream i, by the natural logarithm of the total amount spent by the viewers in those 20 seconds on the aforementioned actions. Considering the lagged effect of viewer tips relative to the broadcaster's performance, we lagged the viewer tipping data by one period.

Lastly, we merged streamers' acoustic emotions, and viewers' danmaku emotions into the consumption data for each 20-second streaming frame. Streaming frames without streamers' speech were excluded. Overall, 5952 observations were retained for analysis. Table 2 and Table 3 report the descriptive statistics of the variables and the correlation matrix.

Table 2. Descriptive Statistics of Variables

Variable	*Observation*	*Mean*	*SD*	*Min*	*Max*
lnTipping	5952	1.275	0.345	0	2.134
Virtual_Streamer	5952	0.463	0.499	0	1
Streamer_Happiness	5952	0.329	0.220	0.00200	0.993
Streamer_Sadness	5952	0.155	0.135	0.00100	0.953
Streamer_Anger	5952	0.166	0.102	0	0.683
Streamer_Neutrality	5952	0.278	0.152	0	0.787
Streamer_Surprise	5952	0.108	0.0980	0	0.600
Streamer_Fearer	5952	0.0250	0.0350	0	0.337
Danmaku _Happiness	5952	0.569	0.288	0.00100	1
Danmaku _Sadness	5952	0.530	0.294	0.00100	1
Danmaku_Anger	5952	0.524	0.297	0	1

Danmaku _Neutrality	5952	0.532	0.294	0	1
Danmaku _Surprise	5952	0.518	0.296	0	1
Danmaku _Fearer	5952	0.515	0.294	0	1
Emotional Synchronicity	5952	0.700	0.129	0.197	0.986

Table 3. Correlation Matrix

	lnTipping	*Virtual _Streamer*	*Emotional Synchronicity*	*Streamer _Happiness*	*Streamer _Sadness*
lnTipping	1				
Virtual_ Streamer	0.604*	1			
EmotionalSynchronicity	0.267*	-0.003	1		
Streamer _Happiness	0.218*	0.233*	0.0023	1	
Streamer _Sadness	-0.002	-0.022	0.105*	-0.418*	1

Note: * p<0.05

4. Empirical result

4.1. The impact of streamer's emotions on streamer-viewer emotional synchronicity

We applied a streamer and time fixed-effect model to explore the relationships between Emotional Synchronicity and streams' emotions. The general model is as follows:

$$Emotional\ Synchronicity_{it} = \lambda_i + \gamma_t + \beta_1\ Streamer\_Emotions + \mu_{it} \quad (1)$$

where the $Emotional\ Synchronicity_{it}$ the cosine similarity between the streamer's voice emotion vector at interval *t* and the viewers' danmaku text emotion vector at interval *t*+1. λ_i is the live-fixed effect, γ_t is the time-fixed effect, and μ_{it} is the error term.

Table 4 reports the effects of the streamer's emotions on Streamer-viewer emotional synchronicity, as illustrated in Equation (1). Firstly, we investigated the impact of the streamer's happiness on the streamer-viewer emotional synchronicity. In column (1) of Table 4, the coefficient for *Streamer_Happiness* is significantly positive, supporting Hypothesis 1a. In column (2) of Table 4, the coefficient for *Streamer_Sadness* is also significantly positive. Therefore, Hypothesis 1b is supported. In column (3) of Table 4, coefficients for both *Streamer_Happiness* and *Streamer_Sadness* are positively significant, further validating Hypotheses 1a and 1b.

Table 4. Results of the Streamer's Emotions on Streamer-viewer Emotional Synchronicity

	(1)	(2)	(3)
	Emotional Synchronicity	*Emotional Synchronicity*	*Emotional Synchronicity*
Streamer_Happiness	0.046***		0.080***
Streamer_Sadness		0.095***	0.143***
Control	Yes	Yes	Yes
Constant	0.686***	0.686***	0.653***
Streamer_FE	Yes	Yes	Yes
Time_FE	Yes	Yes	Yes
N	5952	5952	5952
R^2	0.152	0.155	0.166
adj. R^2	0.077	0.081	0.092

Note: * p<0.1, ** p<0.05, *** p<0.01

4.2. Mediating effect of emotional synchronicity

To examine whether the streamer's emotions can motivate user tipping by promoting emotional synchronicity with the audience, we regard Streamer-viewer emotional synchronicity as a potential mediating variable. Specifically, we hypothesize that the streamer's emotions might influence the streamer-viewer emotional synchronicity, which in turn further affects users' tipping behavior. To validate this mediation effect, we constructed two panel regression models, as depicted in Equation (2) and Equation (3).

$$lnTipping_{it} = \lambda_i + \gamma_t + \theta_1\ Streamer_{Emotion_{it}} + \mu\ Control_{it} + \epsilon_{it} \quad (2)$$

$$lnTipping_{it} = \lambda_i + \gamma_t + \omega_1 Emotional\ Synchronicity_{it} + \omega_2 Streamer\_Emotion_{it} + \mu\ Control_{it} + \epsilon_{it} \quad (3)$$

Table 5 presents the regression results on the impact of the streamer's emotions and their emotional synchronicity with the audience on user tipping behaviors. Columns (1)-(3) of Table 5 report the direct effects of the streamer's emotions on user tipping. Both *Streamer_Happiness* and *Streamer_Sadness* coefficients are significantly positive, indicating that the levels of happiness and sadness displayed by the streamer can motivate users to tip. Furthermore, in columns (4)-(6) of Table 5, we examine the mediating effect of emotional synchronicity. The coefficients for Emotional Synchronicity are 0.740, 0.747, and 0.733 respectively, all of which are significantly positive. This suggests that emotional synchronicity has a significant positive effect on user tipping. Specifically, when a streamer exhibits stronger feelings of happiness and sadness, it can enhance emotional synchronicity with the viewers, subsequently increasing their propensity to tip.

Table 5. The Mediating Effect of Emotional Synchronicity

	(1)	(2)	(3)	(4)	(5)	(6)
	lnTipping	*lnTipping*	*lnTipping*	*lnTipping*	*lnTipping*	*lnTipping*
Emotional Synchronicity				0.740***	0.747***	0.733***
Streamer_Happiness	0.110***		0.152***	0.076***		0.093***
Streamer_Sadness		0.083*	0.174***		0.012	0.069*
Control	Yes	Yes	Yes	Yes	Yes	Yes
Constant	1.237***	1.260***	1.196***	0.729***	0.747***	0.718***
Streamer_FE	Yes	Yes	Yes	Yes	Yes	Yes
Time_FE	Yes	Yes	Yes	Yes	Yes	Yes
N	5952	5952	5952	5952	5952	5952
R^2	0.545	0.542	0.548	0.609	0.608	0.610
adj. R^2	0.505	0.502	0.507	0.575	0.573	0.575

Note: * p<0.1, ** p<0.05, *** p<0.01

4.3. The moderating effect of virtual streamer's emotions on audience tipping behavior

To further investigate whether there is a difference between virtual streamers and human streamers in motivating audience tipping, we introduced a dummy variable describing the type of streamer, termed *Virtual_Streamer*, and interacted it with streamer emotion. Based on the model of Equation (2), we constructed Equation (4) to verify whether the type of streamer plays a moderating role in the impact of streamer emotions on audience tipping.

As previously demonstrated, both the happiness and sadness emotions of the streamer can promote audience tipping by enhancing emotional synchronicity with the audience. If there is a difference in the influence of emotions on audience tipping between virtual and human streamers, there might be two mechanisms leading to this discrepancy: (1) The type of streamer might moderate the effect of the streamer's emotion on emotional synchronicity with the audience. (2) The type of streamer could also potentially moderate the impact of emotional synchronicity on audience tipping.

To identify these two possible mechanisms, we first employed the interaction terms *Virtual_Streamer_Happiness* and *Virtual_Streamer_Sadness* to construct Equation (5), aiming to explore whether there is a differentiated effect between virtual and human streamers concerning emotion and audience emotional synchronicity. Subsequently, using the interaction term with viewer-streameremotional synchronicity, *Virtual_Emotional Synchronicity*, we built Equation (6) to examine whether the pathway from emotional synchronicity to audience tipping differs for virtual streamers.

$$lnTipping_{it} = \lambda_i + \gamma_t + \theta_1 Virtual\_Streamer\_Emotion_{it} + \theta_2 Streamer\_Emotion_{it} + \mu Control_{it} + \epsilon_{it} \quad (4)$$

$$Emotional\ Synchronicity_{it} = \lambda_i + \gamma_t + \theta_1 Virtual\_Streamer\_Emotion_{it} + \theta_2 Streamer\_Emotion_{it} + \mu Control_{it} + \epsilon_{it} \quad (5)$$

$$lnTipping_{it} = \lambda_i + \gamma_t + \theta_1 Virtual\_Streamer\_Emotion_{it} + \theta_2 Streamer\_Emotion_{it} + \theta_3 Emotional\ Synchronicity_{it} + \mu Control_{it} + \epsilon_{it} \quad (6)$$

In Table 6, we primarily investigate the differential impacts of emotions exhibited by virtual streamers as compared to real-human streamers on viewers' tipping behavior.

In column (1), when considering the happiness emotion of virtual streamers, we observe that the coefficient for *Virtual_Streamer_Happiness* is significant and positive (coefficient = 0.157, p<0.01). This suggests that, compared to live human streamers, when virtual streamers display happiness, the tipping behavior of viewers increases. In column (2), the relationship between the sadness emotion of virtual streamers, denoted as *Virtual_Streamer_Sadness* and viewers' tipping is found to be significant and negative (coefficient = -0.360, p<0.01). This indicates that, in comparison to live human streamers, when virtual streamers exhibit sadness, viewers' propensity to tip might decline. Column (3) further synthesizes the differential effects of emotions displayed by virtual streamers on viewer tipping, in comparison to their real-human counterparts.

Overall, Table 6 reveals a notable association between the effects of the streamer's emotions of happiness and sadness on viewers' tipping behavior and the type of streamer. Specifically, relative to real-human streamers, viewers are more inclined to tip when virtual streamers show happiness. Conversely, when displaying sadness, viewers tend to tip real human streamers more.

Table 6. Results of the Moderating Effect

	(1)	(2)	(3)
	lnTipping	*lnTipping*	*lnTipping*
Virtual_Streamer_Happiness	0.157***		0.110***
Virtual_Streamer_Sadness		-0.360***	-0.259***
Streamer_Happiness	0.020		0.082***
Streamer_Sadness		0.280***	0.318***
Control	Yes	Yes	Yes
Constant	1.239***	1.255***	1.195***
Streamer_FE	Yes	Yes	Yes
Time_FE	Yes	Yes	Yes
N	5952	5952	5952
R^2	0.546	0.546	0.551
adj. R^2	0.506	0.506	0.511

Note: * p<0.1, ** p<0.05, *** p<0.01

In Table 7, we further explore the mediating mechanisms of the differentiated effects of virtual streamers compared to real-human streamers on emotions and audience tipping behaviors. In column (1), the relationship between a virtual streamer's happiness emotion and emotional synchronicity is significantly positive, suggesting that virtual streamers enhance the audience's empathy towards the streamer's happiness. However, the relationship between a virtual streamer's sadness emotion and emotional synchronicity is significantly negative, indicating that virtual streamers find it more challenging to evoke audience empathy when expressing sadness compared to real-human streamers.

Column (2) further shows how the emotions of virtual streamers and the emotional synchronicity influence the viewers' tipping behaviors. There's a significantly positive correlation between emotional synchronicity and audience tipping behaviors. Columns (1) and (2) together indicate that the emotions of virtual streamers can increase tipping through emotional synchronicity with the audience. Still, compared to real-human streamers, virtual streamers more readily gain

audience empathy when expressing happiness but struggle more to do so when conveying sadness. These empirical results support Hypotheses 3a and 3b.

Column (3) further tests whether the type of streamer affects the impact of emotional synchronicity on audience tipping, thus explaining the differences between virtual and real-human streamers concerning the influence of emotions on audience tipping. However, the coefficient for *Virtual_Emotional_Synchronicity* is not statistically significant, suggesting that virtual streamers do not significantly alter the degree to which emotional synchronicity affects audience tipping.

Table 7. Additional Tests

	(1)	(2)	(3)
	Emotional Synchronicity	*lnTipping*	*lnTipping*
Emotional Synchronicity		0.724***	0.714***
Virtual_Emotional Synchronicity			0.064
Virtual_Streamer_Happiness	0.073***	0.057*	
Virtual_Streamer_Sadness	-0.087***	-0.196***	
Streamer_Happiness	0.037**	0.056**	0.092***
Streamer_Sadness	0.194***	0.178***	0.070**
Control	Yes	Yes	Yes
Constant	0.652***	0.723***	0.747***
Streamer_FE	Yes	Yes	Yes
Time_FE	Yes	Yes	Yes
N	5952	5952	5952
R^2	0.171	0.611	0.609
adj. R^2	0.097	0.577	0.575

Note: * p<0.1, ** p<0.05, *** p<0.01

Overall, these results provide strong evidence supporting Hypothesis 3, that the emotions of virtual streamers and real-human streamers have varied effects on audience tipping behaviors. The primary mechanism is that compared to real-human streamers, virtual streamers' emotions evoke varying degrees of streamer-viewer emotional synchronicity. Specifically, virtual streamers positively moderate the relationship between the streamer's happiness emotion and emotional synchronicity but negatively moderate the relationship between the streamer's sadness emotion and emotional synchronicity, ultimately influencing the number of tips from the audience.

In the robustness check, we employed an alternative measurement of emotional synchronicity by the ML-based Empathy Dialogue Recognition as a substitute for the original cosine similarity of sadness emotions. Specifically, we utilized a pre-annotated Chinese empathy dialogue dataset provided by Xiaomi. This dataset offers an array of examples, unveiling how individuals reciprocate sympathetically when faced with sadness. We have also adopted an Instrumental Variable (IV) approach, using the frequency of ACG (Anime, Comics, and Games) related vocabulary used by streamers during their live sessions as our instrument. This choice is premised on two assumptions. Firstly, it is plausible to assume that streamers using a higher frequency of ACG-related words are more likely to adopt virtual live streaming, given the cultural and technological overlaps between ACG subcultures and virtual technologies. This satisfies the relevance criterion for an instrumental variable. Secondly, there is no theoretical justification to suggest that merely mentioning more ACG words would directly influence the tipping behavior of viewers. This ensures the exogeneity criterion of the instrumental variable. All of these results were consistent with our prior findings, affirming the robustness of our conclusions. Additionally, we have also identified emotions based on both streamers' facial expressions and included these variables in the models. Due to the page limit, we do not present these results in the manuscript but will provide them upon request.

5. Conclusion and discussion

Live streaming has become increasingly popular in recent years for users to broadcast live events or experiences to an online audience in real-time. With the prevalence of the metaverse, virtual idols appear on the stage of live streaming. Unlike real human streamers, virtual idols are computer-generated avatars whose appearance does not exist in the physical world. Though existing literature has documented the effect of emotional displays of human streamers on viewers' behavior, whether emotional displays of avatar streamers affect viewers' behavior, especially viewers' consumption behavior has not yet been explored. To this end, we focus on the live streaming of virtual idols. Furthermore, we conduct state-of-the-art machine learning algorithms to investigate the emotional synchronicity between streamers and their viewers based on moment-to-moment data. We compare the relationship between steamer-viewer emotional synchronicity and viewers' consumption across the live stream by virtual idols vs. real humans. Our analysis of large-scale video data shows that compared to real-human streamers, virtual streamers are more likely to trigger steamer-viewer emotional synchronicity when expressing happiness, leading to more consumption; but virtual streamers struggle more to trigger steamer-viewer emotional synchronicity when conveying sadness, leading to reduced consumption. In this sense, this study may shed light on how avatars can be designed to leverage emotion to engage viewers in more consumption in live streaming and other marketing contexts. Results from the comparison between live streaming by virtual idols and that by real humans also suggest that in live streaming, especially for those providing entertaining products and services (e.g., singing, dancing, chatting), the mask of animated avatars may induce viewers to consume more only when

streamers express happiness, but real humans can also leverage sadness emotion to boost empathy in viewers and increase their consumption.

The findings of this paper on virtual idol livestreaming and emotional synchronicity have several implications for the livestreaming industry and its stakeholders. First, live-streaming platforms should invest in developing and optimizing features that enhance emotional contagion between virtual idol streamers and their audiences. This could include designing interactive tools and functionalities that facilitate and amplify positive emotions, such as happy interactions. Additionally, platforms can explore ways to enhance the visual and audio effects to create a more immersive experience for the audience. Second, livestreaming companies should consider incorporating virtual idol streamers into their repertoire, as they have shown to be more powerful in eliciting happy emotion contagion. This could involve recruiting, training, and promoting virtual streamers with engaging personalities and personas that resonate with the audience. Third, policymakers should actively engage with the live-streaming industry and contribute to shaping regulations and guidelines that ensure fair competition and protect the interests of both virtual and human streamers. It is important to support innovation and provide a level playing field for all streamers, while also addressing any potential concerns related to emotional exploitation or manipulation. Policymakers should collaborate with industry experts to establish ethical standards that safeguard audience well-being while fostering equitable competition.

References

Aarts, Henk, Ruud Custers, and Hans Marien. 2008. "Preparing and motivating behavior outside of awareness." *Science* (319:5870), pp.1639-1639.

Baevski, A., Zhou, Y., Mohamed, A., & Auli, M. 2020. "wav2vec 2.0: A framework for self-supervised learning of speech representations," *Advances in neural information processing systems* (33), pp.12449-12460.

Batson, C. Daniel, et al. 1991. "Empathic joy and the empathy-altruism hypothesis." *Journal of personality and social psychology* (61:3), pp.413.

Bharadwaj, N., Ballings, M., Naik, P. A., Moore, M., and Arat, M. M. 2022. "A New Livestream Retail Analytics Framework to Assess the Sales Impact of Emotional Displays," *Journal of Marketing* (86:1), pp. 27–47. (https://doi.org/10.1177/00222429211013042).

Brakus, J. Joško, Bernd H. Schmitt, and Lia Zarantonello. 2009. "Brand experience: what is it? How is it measured? Does it affect loyalty?," *Journal of marketing* (73:3), pp. 52-68.

Byrne, Richard W., and Anne E. Russon. 1998. "Learning by imitation: A hierarchical approach," *Behavioral and brain sciences* (21:5), pp.667-684.

Chartrand, Tanya L., and John A. Bargh. 1999. "The chameleon effect: The perception–behavior link and social interaction," *Journal of personality and social psychology* (76:6), pp. 893.

Dwivedi, Y. K. et al. 2022. "Metaverse beyond the Hype: Multidisciplinary Perspectives on Emerging Challenges, Opportunities, and Agenda for Research, Practice and Policy," *International Journal of Information Management* (66:July), Elsevier Ltd, p. 102542. (https://doi.org/10.1016/j.ijinfomgt.2022.102542).

Eckler, Petya, and Paul Bolls. 2011. "Spreading the virus: Emotional tone of viral advertising and its effect on forwarding intentions and attitudes," *Journal of interactive advertising* (11:2). pp. 1-11.

Khwaja, M. G., Jusoh, A., & Nor, K. M. 2019. "Does online social presence lead to purchase intentions?," *International Journal of Economic Policy in Emerging Economies* (12:2), pp.198-206.

Lang, Peter J., and Margaret M. Bradley. 2010. "Emotion and the motivational brain," *Biological psychology* (84:3), pp. 437-450.Lawler, E. J., and Thye, S. R. 1999. "Bringing Emotions into Social Exchange Theory," *Annual Review of Sociology* (25:1), Annual Reviews 4139 El Camino Way, PO Box 10139, Palo Alto, CA 94303-0139, USA, pp. 217–244.

Lee, Y., and Chen, A. N. K. 2011. "Usability Design and Psychological Ownership of a Virtual World," *Journal of Management Information Systems* (28:3), Taylor & Francis, pp. 269–308.

Lin, Y., Yao, D., and Chen, X. 2021. "Happiness Begets Money: Emotion and Engagement in Live Streaming," *Journal of Marketing Research* (58:3), pp. 417–438. (https://doi.org/10.1177/00222437211002477).

Liu, X., Shi, S. W., Teixeira, T., and Wedel, M. 2018. "Video Content Marketing: The Making of Clips," *Journal of Marketing* (82:4), SAGE Publications Sage CA: Los Angeles, CA, pp. 86–101.

Lu, Baozhou, Weiguo Fan, and Mi Zhou. 2016. "Social presence, trust, and social commerce purchase intention: An empirical research," *Computers in human behavior* (56), pp. 225-237.

Teixeira, T., Wedel, M., and Pieters, R. 2012. "Emotion-Induced Engagement in Internet Video Advertisements," *Journal of Marketing Research* (49:2), SAGE Publications Sage CA: Los Angeles, CA, pp. 144–159.

Wang, X., Jiang, M., Han, W., & Qiu, L. 2022. "Do emotions sell? The impact of emotional expressions on sales in the space-sharing economy," *Production and Operations Management* (31:1), pp.65-82.

Wohn, D. Y., and Freeman, G. 2020. "Live Streaming, Playing, and Money Spending Behaviors in ESports," *Games and Culture* (15:1), SAGE Publications Sage CA: Los Angeles, CA, pp. 73–88.

Xu, Y., Ye, Y., and Liu, Y. 2022. "Understanding Virtual Gifting in Live Streaming by the Theory of Planned Behavior," *Human Behavior and Emerging Technologies* (2022), pp. 1–12. (https://doi.org/10.1155/2022/8148077).

Ye, Shun, et al. 2020. "Social presence, telepresence and customers' intention to purchase online peer-to-peer

accommodation: A mediating model," *Journal of Hospitality and Tourism Management* (42), pp.119-129.

Zhang, Zhengyan, et al. 2019. "ERNIE: Enhanced language representation with informative entities." *arXiv preprint arXiv:1905.07129*.

Zhang, Q., Wang, W., and Chen, Y. 2020. "Frontiers: In-Consumption Social Listening with Moment-to-Moment Unstructured Data: The Case of Movie Appreciation and Live Comments," *Marketing Science* (39:2), pp. 285–295. (https://doi.org/10.1287/mksc.2019.1215).