Journal of Business Research 208 (2026) 116050

Contents lists available at ScienceDirect

Journal of Business Research

journal homepage: www.elsevier.com/locate/jbusres

Face reading technology: improving preference prediction from self-reports using micro-expressions

Pantelis Karapanagiotis [a], Franziska Krause [b,*], Janina Krick [c]

[a] Postdoctoral Researcher of Microeconomics, University of Groningen, Broerstraat 5, 9712 CP Groningen, Netherlands; Leibniz Institute for Financial Research SAFE, Goethe University Frankfurt, Germany

[b] Associate Professor of Marketing & Customer Insight, EBS University of Business and Law, Rheingaustrasse 1, 65375 Oestrich-Winkel, Germany

[c] Doctoral Student in Marketing, EBS University of Business and Law, Rheingaustrasse 1, 65375 Oestrich-Winkel, Germany

ARTICLE INFO

Keywords:
Choice modeling
Face reading technology
Micro-expressions
Machine learning
Preference prediction

ABSTRACT

Traditional market research primarily relies on self-reports (SR) to assess consumer preferences, yet these methods are prone to biases and limited in capturing subconscious emotional responses. Psychophysiological and neurophysiological methods offer objective alternatives, but their high cost and intrusiveness limit practical use. This study examines automated Facial Expression Analysis (FEA), focusing on micro-expression (ME) emotion data, as a scalable, non-intrusive approach to improve the accuracy of predicting consumer choices. In a controlled experiment exposing participants to both video and poster ads, we compare the predictive power of ME and SR emotion data using machine learning and artificial neural network models. Results demonstrate that ME data significantly enhance both multinomial and binomial choice prediction accuracy compared to SR data, particularly for dynamic video ads where ME patterns capture real-time emotional fluctuations more effectively. Beyond its methodological contributions, our study underscores practical implications and discusses the ethical considerations related to consumer privacy.

1. Introduction

Consumer preference elicitation and choice predictions are fundamental for effective marketing strategies (Dixon & Mikolon, 2021; Ghiassaleh et al., 2020; Kushwah et al., 2019; Wichmann et al., 2022). Seminal studies highlight the critical role of emotions in shaping consumer preferences and decision-making (Ekman, 1992; Ekman & Friesen, 1971; Lerner et al., 2015). Traditional market research relies on self-reports, including questionnaires, focus groups, and interviews, to collect this information (Ozretic-Dosen et al., 2022; Wilkinson & Birmingham, 2011). Despite their widespread use, self-reports are often criticized for their susceptibility to cognitive biases (MacKenzie & Podsakoff, 2012; Neeley & Cronley, 2004), sensitivity to framing (Buchanan & Henderson, 1992; Griffin & Hauser, 1993; McDaniel & Kolari, 1987; Shephard, 2003), inconsistencies in measurement scales (Holbrook & Batra, 1988), and an inability to capture subconscious emotional responses (Johansson et al., 2006; Richins, 1997).

Psychophysiological and neurophysiological methods, providing objective measurements of emotional states, have been employed to overcome the limitations of self-reports (Bazzani et al., 2020; Blasco-Arcas et al., 2022; Venkatraman et al., 2015; Yoon et al., 2009). However, these methods require costly specialized equipment, are intrusive, and lack scalability in both research and practical applications (Ariely & Berns, 2010; Camerer & Yoon, 2015; Plassmann et al., 2012). Automated Facial Expression Analysis (FEA) presents a non-intrusive, scalable alternative for objectively measuring emotional engagement by capturing micro-expressions (ME) (Kulke et al., 2020); brief, either conscious or subconscious, facial cues that can reliably serve as indicators of genuine emotions (Den Uyl & Van Kuilenburg, 2005).

Recent literature in consumer research demonstrates FEA's potential for explaining video ad attitude and effectiveness (Höfling et al., 2023; Lewinski et al., 2014; Li et al., 2018). Nevertheless, while a series of studies have established that psychophysiological and neurophysiological measurements can improve prediction accuracy of consumer choices (Hakim et al., 2021; Kong et al., 2013; Ravaja et al., 2013; Telpaz et al., 2015; Yadava et al., 2017), remarkably little is known about whether automated FEA measurements lead to similar improvements. Furthermore, prior research on facial expression analysis (FEA) has predominantly examined video advertising contexts. However, advertising formats differ substantially in how consumers are exposed to

* Corresponding author.
E-mail addresses: p.karapanagiotis@rug.nl (P. Karapanagiotis), franziska.krause@ebs.edu (F. Krause), janina.krick@ebs.edu (J. Krick).

https://doi.org/10.1016/j.jbusres.2026.116050
Received 6 December 2023; Received in revised form 3 February 2026; Accepted 3 February 2026
Available online 13 February 2026

and process marketing stimuli. In practice, firms continue to rely heavily not only on video advertising but also on poster-based formats such as digital display, out-of-home, and traditional print (e.g., Du et al., 2019; Bayer et al., 2020). This raises the question of whether and how the effectiveness of FEA generalizes across different types of ad exposure.

Our study provides novel evidence on the impact of ME emotion measurements from FEA on choice prediction accuracy. We compare the contributions of self-reported (SR) and ME emotions in multinomial (one out of many) and binomial (one out of two) choice predictions. We investigate these contributions both when emotion measurements are used as standalone predictors and when they are combined with additional survey measurements of ad and product characteristics. We find that emotions, either measured by SR or ME, can improve choice prediction accuracy compared to exclusively using ad and product survey measurements. Notably, using ME data yields greater prediction accuracy than SR on average. Nonetheless, our analysis corroborates that these improvements are heterogeneous across types of exposures. We find that the effectiveness of ME data varies by ad format, with stronger predictive accuracy observed for video ads compared to poster ads.

The study employs a controlled experimental design involving 103 participants exposed to either video or poster ads for confectionery products. We collected data encompassing participant demographics, survey responses on ad and product characteristics, and both SR and ME emotion measurements. ME data was extracted from visual recordings of participants' faces during ad exposure using the FaceReader tool. Unlike SR data, which offer a single, retrospective measure, ME data provides a real-time, dynamic representation of emotional responses during ad exposure.

A key technical focus of our analyses is the leverage of the dynamic nature of ME data. To analyze this data, we used a comprehensive suite of Machine Learning (ML) and Artificial Neural Network (ANN) methods in both the multinomial and binomial choice settings, allowing us to evaluate the predictive power of ME data in different choice scenarios. Our ANN approach offers a cost-effective, non-intrusive, and scalable alternative to existing resource-intensive psycho- and neurophysiological methods employed in choice prediction research for incorporating non-SR emotion measurements.

The results of our analyses indicate that combinations of automated FEA measurements and predictive ANN models have the potential to find broad applications in research and practice. Mainly, we highlight the benefits of integrating such technologies into ad design processes. By leveraging real-time ME data, marketers can refine ad designs more dynamically, moving beyond traditional A/B testing to enable faster and more adaptive design-test-learn cycles. Particularly for video ads, FEA allows continuous emotion monitoring throughout exposure, providing in-depth insights into consumer preferences. When paired with predictive ANN models, this granular data can increase the accuracy of consumer choice predictions, supporting more effective ad targeting, higher conversion rates, and more efficient allocation of marketing resources (Bayer et al., 2020; Brandt, 2016; Garaus et al., 2021).

At the same time, the use of FEA measurements with predictive ANN models for preference elicitation raises ethical and privacy concerns. We outline these concerns and discuss the implications of our findings in the context of consumer privacy and AI safety.

2. Literature Review

2.1. Emotional measurements in marketing research

Basic emotions—such as happiness, sadness, anger, fear, surprise, disgust, and contempt—are universal, and their associated facial expressions are recognizable across cultures (Ekman & Friesen, 1971; Ekman, 1992). These emotions structure how consumers perceive and respond to marketing stimuli at both conscious and subconscious levels (Bagozzi et al., 1999; Dimberg et al., 2000; Tuan Pham, 2004). Also, emotions are central to core marketing outcomes: they guide assessments of product attributes, shape risk perceptions, and foster brand loyalty (Keller, 2012; Lerner et al., 2015). Illustratively, positive emotions (e.g., happiness, excitement) tend to enhance product evaluations and increase purchase intentions, whereas negative emotions (e.g., fear, sadness) often narrow consideration sets and trigger avoidance behaviors (Donovan et al., 1994; Wang et al., 2011).

Accurate measurement of consumer emotions is crucial in marketing research. However, traditional self-reported methods have several limitations. While widely used, self-reports are often criticized for their susceptibility to cognitive biases, lack of standardization, inconsistencies in measurement scales, and an inability to capture subconscious emotional responses (Holbrook & Batra, 1988; Richins, 1997). To capture these dynamics effectively, this work emphasizes accurate, granular measurement approaches that register discrete emotional episodes and their temporal evolution during ad exposure. Section 1.2 reviews such non-self-reported methods and situates automated Facial Expression Analysis within this toolkit.

2.2. Non-self-reported emotional measurements

An increasing number of advertising studies explore modern technologies that measure emotions objectively, addressing the limitations of self-reports (Poels & Dewitte, 2006). Hazlett and Hazlett (1999) found that facial EMG measurements outperform self-reports in distinguishing between video ads and predicting ad recall. Similarly, Li et al. (2018) compared self-reports with psychophysiological measures (facial EMG and skin conductance) in tourism advertising and found that while self-reports provided useful insights, they often failed to reflect emotional responses accurately. Venkatraman et al. (2015) demonstrated that neurophysiological methods, such as fMRI and EEG, can more effectively link individuals' reactions to television ads with aggregate market outcomes compared to self-reports.

Despite these advantages, psychophysiological and neurophysiological methods require specialized equipment and are often intrusive, costly, and difficult to scale for large consumer studies (Ariely & Berns, 2010; Camerer & Yoon, 2015; Plassmann et al., 2012). FEA offers a non-intrusive, scalable alternative for objective measurement of emotional engagement by detecting micro-expressions—brief, involuntary facial muscle activations operationalized as changes in facial action units (Kulke et al., 2020; Den Uyl & Van Kuilenburg, 2005; Lewinski et al., 2014).

The analysis of nonverbal facial cues has a long-standing history in academic research. Darwin (1965/1872) provided detailed descriptions of human facial expressions, arguing that these expressions are universal and innate, playing an essential role in communication. This early work laid the foundation for the development of the Facial Action Coding System (FACS) by Ekman and Friesen (1978). FACS classifies facial expressions using 46 action units, each linked to a specific muscle movement (e.g., raising or lowering the corners of the mouth).

Early research relied on manual coding of facial expressions using FACS, a time-consuming process requiring trained observers (Derbaix & Bree, 1997). This method limited large-scale studies and raised concerns about inter-coder consistency, particularly when interpreting subtle muscle movements (Derbaix, 1995). Recent advances in ML have enabled AI tools to automate ME coding (Bartlett et al., 2006; Cohn & De La Torre, 2015; McDuff et al., 2013; Tian et al., 2001), facilitating research across various decision-making contexts (Masip et al., 2014; Sato et al., 2022; Tkalčič et al., 2019).

Compared to other objective emotional measurement methods, automatic FEA is non-invasive, scalable, and highly applicable in real-world marketing. For example, companies like Procter & Gamble, Unilever, and GfK use FEA during ad design and testing to assess engagement and optimize advertising effectiveness (Teixeira et al., 2012). Therefore, prior advertising research has focused more strongly on the relationship between FEA emotions and ad effectiveness for dynamic stimuli (e.g., video, real-time signage), rather than for static formats, with most studies additionally relying on categorical emotion outputs rather than time-evolving action unit features. These patterns, together with limited direct benchmarks against self-reports for choice prediction, motivate our focus on granular, dynamic ME measurements and format-specific moderation.

2.3. Automatic FEA tools

Common AI tools for automatic FEA in consumer research include FaceReader by Noldus (https://www.noldus.com), Affdex by Affectiva (https://www.affectiva.com), and Cognitive Services by Microsoft (https://azure.microsoft.com/de-de/products/cognitive-services). These tools rely on standard (web) cameras, are designed based on FACS, and produce comparable results (Dupŕe et al., 2020; Skiendziel et al., 2019).

Our analysis uses emotion measurements collected by FaceReader (version 9; technical notes documented by Loijens and Krips (2025)). FaceReader identifies 20 facial action units (e.g., eyebrow raises) and detects even brief expressions, associating them with specific emotions. Captured action units are scored and annotated as Trace, Slight, Pronounced, Severe, or Max. These scores are processed by an ANN that generates intensity scores for eight core emotional states: happiness, sadness, anger, fear, surprise, disgust, contempt, and neutrality. For example, neutral emotions are measured by combinations of expressions that are associated neither with positive nor with negative action units (see Gasper et al., 2024, for a discussion of the role of neutral emotional states). Each emotional state is quantified independently on a scale from 0 (no activation) to 1 (maximum activation). Since these states are independently calculated, their sum can exceed one. FaceReader also provides two synthetic emotional states, which represent a composition of different core emotional states and action units: arousal (0 to 1) and valence (−1 to 1).

Arousal is calculated from past and current action unit scores, providing an overarching measure of facial activation. First, the average action scores (AAS) over the last (at most) 60 s are calculated for each action unit. Then, AAS is subtracted from the current action score (AS) to correct the AS for individual differences in facial expressiveness. The corrected action score is then calculated by $CAS = \max(0, AS - AAS)$. Finally, arousal is calculated by the average of the top 5 CAS among the CAS of all 20 action units. Valence is calculated based on the difference between the activation of happiness (as the only unambiguously positive emotional state) and the highest unambiguously negative emotional state (one among sad, angry, scared, and disgusted).

2.4. FEA applications in consumer research

Facial expression analysis (FEA), whether manually coded or automated, has been applied across diverse areas of consumer research. Table 1 groups representative studies into four clusters: foundational work, early advertising applications with manual coding, automated applications in advertising contexts, and extensions beyond advertising. This clustering illustrates how the field has evolved while also revealing commonalities and limitations.

Foundational work established the methodological basis for FEA. Ekman and Friesen (1971) identified universal expressions linked to basic emotions as core of their developed FACS, laying the groundwork for analyzing consumer responses via facial cues. Scherer and Ellgring (2007) extended this by combining FACS with physiological measures, showing systematic associations between expressions and physiological changes. Together, these studies provided the conceptual and methodological foundation for later consumer applications.

Early applications in advertising relied on manual coding and highlighted methodological constraints. Derbaix (1995) showed that facially coded expressions contributed little beyond verbal affective reactions when predicting ad responses, and Derbaix and Bree (1997) reported similar findings in children's ad evaluations. These studies demonstrated the importance of affect for advertising effectiveness but also underscored the limits of manual FEA compared to verbal measures.

Automated applications in advertising contexts demonstrated stronger predictive power. Lewinski et al. (2014) showed that happiness detected via automated FEA predicted attitudes toward amusing ads, though effects were limited to medium- and high-amusement stimuli. Mehta et al. (2021) found that positive facial expressions during energy

Table 1
Summary of FEA in Consumer Research Literature.

Cluster	Study	Objective	Methodology	Key Findings	Contribution to Consumer Choice Research
Foundational work	Ekman & Friesen (1971)	To develop a system for categorizing facial expressions.	Facial Action Coding System (FACS).	Identified universal expressions associated with basic emotions.	Laid the foundation for analyzing consumer emotional responses via facial expressions.
	Scherer & Ellgring (2007)	To investigate facial expressions in emotional contexts.	FACS combined with physiological measures.	Correlation between specific expressions and physiological changes.	Advanced understanding of emotional responses in applied settings.
Manual FEA in advertising contexts	Derbaix (1995)	To assess affective reactions in advertising.	Verbal and facial affective reactions manually coded.	Verbal measures predicted ad responses; facial coding added little.	Highlighted methodological limits of manual FEA and value of multi-method approaches.
	Derbaix & Bree (1997)	To examine children's affective responses to ads.	Experimental study with manual coding and verbal reports.	Verbal measures outperformed facial coding in predicting attitudes.	Showed the role of affect in shaping children's brand and ad evaluations.
Automated FEA in advertising contexts	Lewinski et al. (2014)	To predict ad effectiveness using FEA.	Experimental study of amusing ads with FEA and self-reports.	Happiness predicted attitudes for medium/highly amusing ads.	Demonstrated automated FEA's potential to predict ad outcomes.
	Garaus et al. (2021)	To examine FEA-based digital signage at the point-of-sale.	Lab and online experiments with real-time facial recognition.	Emotional targeting increased purchase intent; disclosure reduced effects.	Illustrated both potential and ethical concerns of FEA-driven targeting.
	Mehta et al. (2021)	To explore FEA in product sampling.	Combined self-reports and FEA during energy drink tasting.	Positive FEA emotions aligned with SR and correlated with acceptability.	Showed convergent validity and potential for predicting product evaluations.
Applications beyond advertising	Kerrihard et al. (2017)	To assess acclimation/demographics in food responses.	FEA and SR during salt tasting.	FEA and SR showed acclimation effects.	Demonstrated how FEA captures individual taste preferences.
	González-Rodŕiguez et al. (2020)	To apply facial recognition in tourism.	Automated facial expression recognition in real-time.	Facial expressions predicted tourist satisfaction and recommendations.	Extended FEA applications to services and tourism.
	De Wijk et al. (2022)	To examine immersive food contexts.	FEA and SR in simulated consumption settings.	FEA more sensitive to subtle affective changes.	Showed context-dependence and potential of FEA in food research.

drink sampling aligned with self-reports and correlated with product acceptability. Garaus et al. (2021) extended this to field-like settings, showing that real-time digital signage using FEA-based targeting could increase purchase intentions, although disclosure of such practices attenuated the effect. Collectively, these studies suggest that automated FEA in dynamic, real-time contexts adds explanatory value beyond self-reports. At the same time, outcomes remain largely confined to attitudinal proxies (e.g., attitudes, sympathy, recall), and most studies emphasize categorical emotion labels such as "happiness" rather than fine-grained or time-resolved facial features.

Applications beyond advertising indicate that FEA can capture meaningful emotional responses in consumption and service contexts while also highlighting strong context dependence. Kerrihard et al. (2017) used FEA and self-reports to show acclimation and demographic influences on salt preferences. González-Rodríguez et al. (2020) demonstrated that facial recognition in real-time tourist experiences predicted satisfaction and recommendation likelihood. De Wijk et al. (2022) further showed that in immersive food consumption contexts, FEA could be more sensitive than self-reports to subtle affective differences. These findings provide convergent validity but also raise questions about the generalizability of FEA effects across domains.

Taken together, the literature illustrates an important trajectory: from methodological foundations and early manual attempts with limited incremental value to automated FEA demonstrating predictive validity in advertising and beyond. While FEA has become more reliable and versatile, its applications remain partially narrow in scope, with a focus on specific emotions and attitudinal outcomes.

2.5. Preference prediction beyond Self-Reports

The role of emotions in predicting consumer preferences has long been recognized as a central factor in understanding consumer behavior. Emotions influence both immediate responses to marketing stimuli and longer-term attitudes toward brands and products. Traditionally, SR measures, such as questionnaires and surveys, have been the primary tool to capture these emotional responses. However, SR measures are inherently limited by participants' introspective abilities and potential biases, such as social desirability or memory distortions. This limitation has prompted an increasing number of studies to explore non-SR measurements, including FEA, physiological recordings, and neuroimaging, to provide a more objective and fine-grained assessment of emotional responses. Table 2 provides an overview of studies that compare the predictive accuracy of SR and non-SR measurements in relation to consumer preferences and brand attitudes.

A first cluster of early studies investigated whether non-SR measures could provide incremental predictive value beyond SR. Derbaix (1995) found that SR remained the stronger predictor of ad and brand attitudes, highlighting the limitations of manual coding for facial expressions and the nascent stage of this research at the time. Hazlett and Hazlett (1999) and Lewinski et al. (2014) extended this inquiry by demonstrating that non-SR measures, particularly FEA, can capture subtle emotional responses that are not always consciously accessible, thereby influencing recall, engagement, and attitude formation. These studies collectively underscore the potential of integrating objective emotional measures to complement traditional survey-based methods.

A second cluster of research leverages multimodal approaches and advanced analytics to deepen our understanding of the predictive role of emotions. Venkatraman et al. (2015) showed that neuroimaging techniques, such as fMRI, can significantly enhance predictions of advertising effectiveness beyond conventional self-reports. Li et al. (2018) demonstrated the complementarity of SR and FEA in shaping destination attitudes, highlighting that different measures may capture distinct emotional dimensions with varying behavioral consequences. They showed that when evaluating advertisements, physiological measures were more sensitive to subtle emotional reactions, while self-reports were influenced by social desirability and retrospective biases. They also linked positive emotional reactions, particularly joy, to increased purchase intentions and more favorable brand perceptions, highlighting

Table 2
Preference Prediction with Non-SR Measurements.

Study	Context	Measures	Stimuli	Methods	Key Findings (Empirical Results)	Added Contribution (vs. prior work)
SR and FEA						
Derbaix (1995)	Familiar & unfamiliar brands	SR emotions; FEA	Video ads	Regression	SR explained ad & brand attitude; FEA added no explanatory power.	Early test of FEA, showed SR as robust predictors, questioned the incremental value of FEA.
Hazlett & Hazlett (1999)	Cars, telecom, food, movies	SR knowledge & relevance; SR emotions; facial EMG	Video ads	ANOVA	Emotionally evocative ads recalled more strongly; EMG captured subtle differences.	Demonstrated added diagnosticity of psychophysiological measures for recall.
Lewinski et al. (2014)	Food, e-trade, cosmetics	Ad & brand attitude; SR + FEA	Video ads	Correlation	Positive FEA linked to ad & brand attitude, esp. in amusing ads.	Provided evidence for FEA's predictive role in attitudinal outcomes.
Multimodal approaches						
Venkatraman et al. (2015)	Consumer goods, services, travel	SR (liking, recall, intention); IAT, eye tracking, EEG, fMRI	Video ads	Regression	fMRI improved prediction of ad elasticities beyond SR.	Established neuro measures as incremental predictors of advertising effects.
Li et al. (2018)	Tourism products	SR (destination attitude); SR emotions; facial EMG	Video ads	PLS	Both SR & EMG explained destination attitudes; SR stronger for conscious attitudes, EMG captured subconscious reactions.	Showed complementarity of SR and psychophysiological measures in predicting market-relevant outcomes.
Hakim et al. (2021)	Food products	ARF Survey; EEG	Video ads	ML classifiers (SVM, Trees, KNN)	Up to + 20 pp prediction improvement vs. SR.	Demonstrated that ML approaches with EEG outperform traditional surveys.
Höfling et al. (2023)	Various industries	Demographics; Anxiety; Expressivity; SR + FEA	Video ads	Semiparametric mixed models	FEA predicted joy, ad likeability, brand likeability, purchase intention.	Validated FEA as consistent predictor across industries.
This Study						
This Study	Confectionery products	ARF Survey; SR emotions; FEA	Video & poster ads	ML classifiers (LDA, SVM, Boosting, ANN, LSTM)	+3–5 pp over SR in video ads; weaker gains for posters; consistent within-sample improvement.	Extends prior work by systematically comparing exposure formats (video vs. poster) and testing DL classifiers, clarifying boundary conditions for FEA's predictive value.

the added value of automated FEA for predicting consumer responses. Höfling et al. (2023) further advanced this line of inquiry by integrating FEA with machine learning algorithms, showing that automated analysis of emotional expressions can reliably predict consumer responses, including joy, ad likeability, and purchase intention, across diverse product categories. These studies collectively illustrate an evolution from descriptive assessments of emotional responses to predictive frameworks that harness technological and analytical advances to enhance marketing insights.

Building on this progression, the present study contributes in two key ways. First, it systematically compares different exposure formats (video vs. poster ads), addressing the gap in previous research that largely focused on video stimuli, thereby testing the robustness of FEA across diverse advertising modalities. Second, it employs a broader set of machine learning approaches, including deep learning classifiers, to evaluate the predictive power of FEA relative to SR measures in forecasting consumer choices and brand attitudes. By integrating these methodological and analytical advancements, the study provides a nuanced understanding of the conditions under which non-SR measures meaningfully augment prediction accuracy, offering both theoretical insights and practical guidance for advertising research.

3. Experimental design and data collection

Our study contributes to the literature investigating preference prediction with non-SR measurements by comparing FEA and SR emotion measurements across different ad types in a controlled experiment. This section details our experimental design and data collection methodology.

3.1. Experiment Description

One hundred three participants (28% female, 68% male, 4% other; Mage = 30.88, SD = 9.71) participated in this study. Fig. 1 visualizes the experimental design. The study comprised three main stages. Table 3 provides an overview of the measurements collected at each stage.

In the first stage of the experiment, participants were randomly assigned to one of two groups. Participants in the first group watched six video ads for confectionery products, each lasting between 15 and 20 s. Those in the second group were shown six poster ads for the same products, each for 3 s.

The video and poster ads were selected from real-world advertising campaigns, ensuring that they accurately reflected how companies market their products, thus enhancing the ecological validity of our analysis. The choice to display each poster ad for 3 s balances our aim to maintain ecological validity with evidence from prior research. In everyday life, consumers glancing at poster ads typically view them for only a few seconds rather than engaging with them for extended periods. More importantly, previous studies on ad effectiveness support the notion of short attention spans for poster ads. For example, Pieters and Wedel (2004) used eye-tracking technology and found that consumers spent an average of 1.73 s looking at a poster ad and 2.79 s looking at editorial pages.

The selection of confectionery products was based on several factors. First, they have broad consumer appeal across various demographics, including age and gender, making them suitable for our study. Additionally, we aimed to measure a real-world take-out choice as a downstream behavior, which required providing actual products to participants. Confectionery items were both affordable and practical for this purpose, as they could be distributed individually. The confectionery had real-world prices that were roughly comparable, ranging from $1.20 to $3.29 (1.09€ to 2.99€). Web Appendix A contains additional product information, including product images and links to the ads used. The total duration of all video ads was 94 s, and the total duration of all poster ads was 18 s.

We captured visual recordings of participants' faces during exposure to the video or poster ads. We extracted each participant's ME emotion measurements during ad exposure from these recordings using the FaceReader tool. The collected measurements and their corresponding scales are detailed in the stage 1 section of Table 3.

In the second stage, after exposure to all six ads, participants completed a brief questionnaire about each product. They reported their emotions for the ads they watched and answered survey questions about the ads and the confectionery products. For each measurement, Table 3 lists the question asked, the source from which the question was derived, and the scale used.

The measured SR emotions were selected to correspond with the ME measurements provided by the FaceReader. The survey questions were adapted from the ARF survey by the Advertising Research Foundation in New York, which is frequently used in market research (Venkatraman et al., 2015). We use these ARF survey questions as a starting point for our analysis and compare the additional choice prediction accuracy obtained from using ME and SR measurements. ARF survey questions were also used by Hakim et al. (2021) as a starting point for

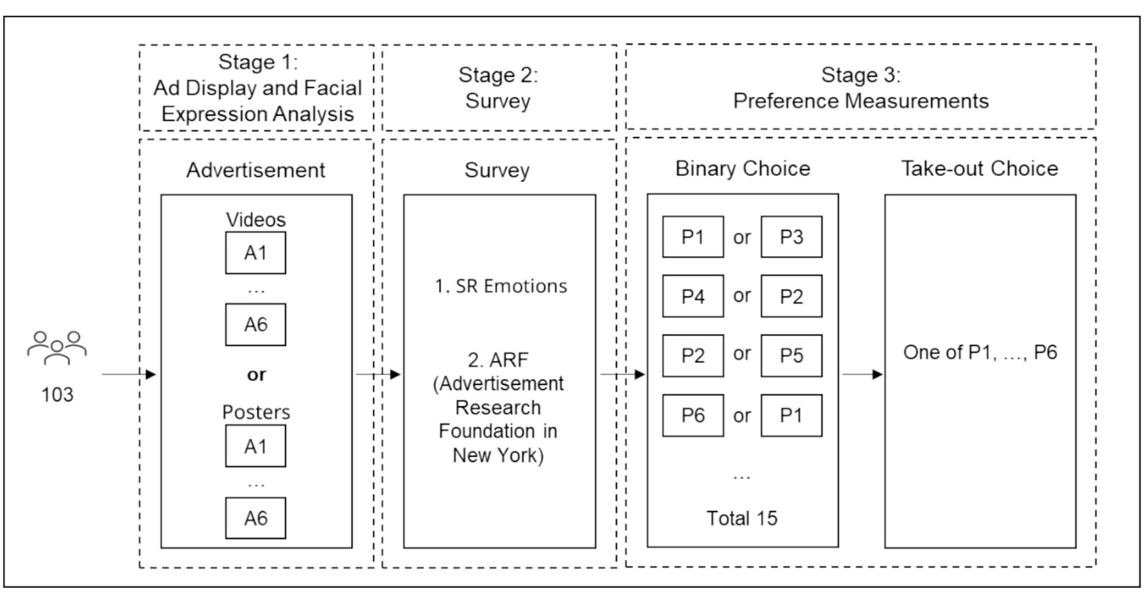


Fig. 1. Experimental Procedure.

Table 3
Description of Measurements.

Measure	Scale	Description	Source
ME Emotions (Stage 1)			
Happy, Sad, Angry, Surprised, Scared, Disgusted, Contempt, Neutral, Arousal	0 to 1	FEA measured	FaceReader (v9) (Loijens & Krips, 2025)
Valence	−1 to 1		
Participants' Characteristics (Stage 2)			
Age	In years		
Gender	Male; Female; Other		
ARF (Stage 2)			
I like it	1 (strongly disagree) to 7 (strongly agree)	I like the product.	Adapted from ARF survey (Haley & Baldinger, 2000)
High quality		The product is of high quality.	
Enjoyable		The product is pleasant.	
Appetizing		The product tastes good.	
Purchase Intention 1		If I come across the product in a store, I will buy it.	
Purchase Intention 2		I will search for and buy the product the next time I am in a store that sells it.	
Purchase Intention 3		I would recommend someone else to buy the product.	
Purchase Intention 4		After seeing the advertisement, I am eager to try the product.	
Product Knowledge	1 (not at all) to 7 (very much)	Do you know the product from the advertisement?	
Product Consumption		How frequently do you consume the advertised product?	
Ad Likability		How much did you like the advertisement?	
Ad Effectiveness		How effective do you find the advertisement in your opinion?	
Familiarity	Yes; I am not sure; No	Have you seen this advertisement before?	
SR Emotions (Stage 2)			
As in Stage 1	1 (not intensely at all) to 7 (very intensely)	Please indicate how intensely you experienced the following emotions while watching the advertisement.	
Preferences (Stage 3)			
Binary Choices	Product choice	One out of two products × 15 combinations	
Take-out Choice		One out of six products	

investigating the additional choice prediction accuracy obtained from using EEG measurements.

In the third stage of the experiment, participant preferences were measured using two independent approaches. First, participants compared all possible product pairs. Each participant was presented with all 15 possible comparisons of the six different products in random order. Each product pair was displayed for a maximum of 3 s. Participants selected a product by clicking on one of the product images. If a participant did not decide within the allotted time, the trial was considered a mistrial and excluded from the analyses. Second, participants chose a product to take with them.

3.2. Data Description

The raw data collected from our experimental procedure includes information on participants, ads, and products. The data file as well as the replication code used for the analysis can be found and accessed on OSF and GitHub (https://osf.io/u3y9j/overview?view_only=5846a80042e3496e9e92cb9ec03496b9;https://github.com/pi-kappa-devel/mex).

Fig. 2 visualizes the interdependencies of the collected data. For each participant, we collected age and gender. For each participant-product combination, we (i) collected survey data on ad and product characteristics and (ii) documented participants' preferences through binomial product comparisons and take-out choices. Finally, for each ad-participant combination, we measured participants' emotions using both SR and ME measurements. ME measurements are dynamic; they were collected during ad exposure in 200-millisecond intervals. SR measurements were collected after ad exposure. From the raw data, we structured the two datasets used in our analysis.

The first dataset focuses on the participants' take-out choices among the six products of the experiment. The experiment's framing effectively fixes the choice sets of the participants to the six products included in our investigation. We use this dataset to examine the participants' multinomial choice problem in this choice set. For each participant and product ($N \times J = 618$), we construct a response variable labeled as success (has a value of 1) if the participant has chosen the product and labeled as failure (has a value of 0) otherwise. The response variable of the first dataset is accompanied by participant, ad and product characteristics, and SR and ME emotion explanatory variables in levels.

The second dataset focuses on participants' preferences as revealed by the binary comparisons of the products. For each participant and any combination of two products, say *Left* and *Right*, we examine the comparison $Left \preceq Right$, and assign the value 1 to the response variable if the participant selects *Right* over *Left* and 0 otherwise. This labeling gives a sample of $N \times P = 1545$ observations. The second dataset has 2.5 times the number of observations and a more balanced response variable than the first dataset. The resulting response variable is equal to 1 between 4 and 12 times for each participant, and participants have, on average, chosen approximately 49% of the time *Right* instead of *Left*, reflecting that the product order in binary comparisons was randomized.

The response variable of the second dataset is accompanied by participant, ad and product characteristics, and SR and ME emotion explanatory variables, expressed as cross-product differences. For example, if x_l, x_r, denote the "Purchase intention" responses of a participant for products *Left* and *Right*, we construct a variable $y = x_r - x_l$ denoting the participant's difference in the purchase intention between *Right* and *Left*. This approach eliminates any participant-specific biases in the survey and FEA measurements.

4. Methods and results

4.1. Statistical methods

We implement a comprehensive collection of ML and ANN methods for both the multinomial and binomial choice data to assess the predictive capacity of ME measurements on product choices. Table 4 presents a summary of the employed methods. We present the results of Linear Discriminant Analysis (LDA), Deep Learning (DL), and Long Short-Term Memory (LSTM) models for the multinomial choice problem. For the binomial choice problem, we present the results of Logit, Boosting, Support Vector Machines (SVM), DL, and LSTM models.

For the static models in our analysis (LDA, Logit, Boosting, SVM, and DL), we follow Hakim et al. (2021) and aggregate the time dimension of our raw data into single measurements for each participant and product. The aggregated emotion features are obtained using two summary statistics. First, average ME emotion variables are calculated as averages over the measurements during the run-time of the ad exposure. Second,

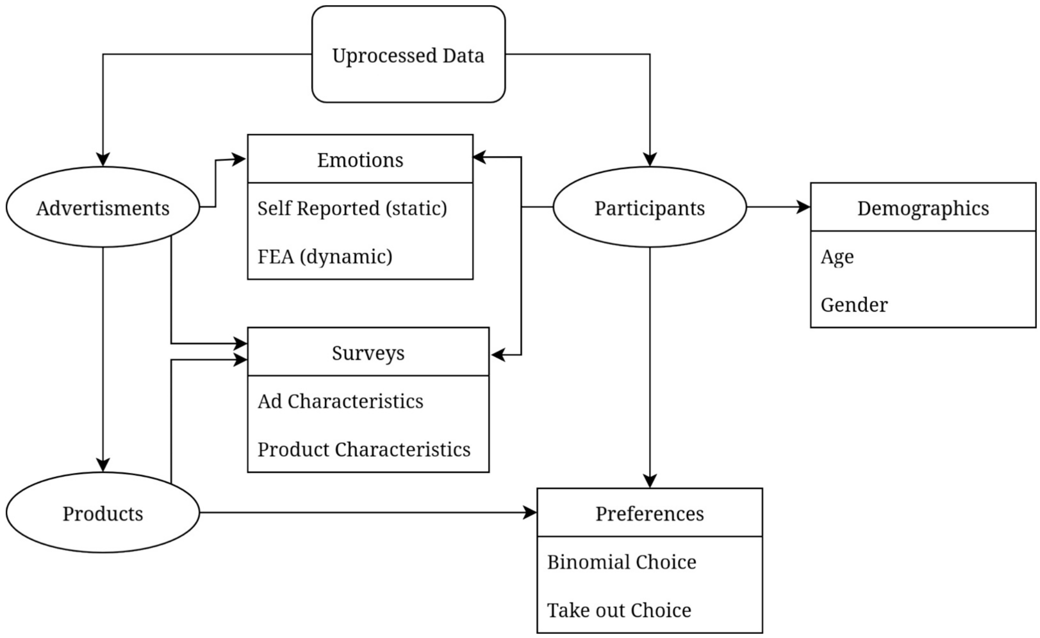


Fig. 2. Collected Data Interdependencies.

Table 4
Statistical Methods.

Method	Time	FEA Aggregation
(a) Multinomial Choice Models.		
LDA	Static	Average
DL	Static	Average
LSTM	Dynamic	None
(b) Binomial Choice Models.		
Logit	Static	Average
Boosting	Static	Average, Extreme
SVM	Static	Average
DL	Static	Average
LSTM	Dynamic	None

extreme ME emotion variables are calculated as the maximum emotion measurement observed for each ad's duration.

The ANN (DL and LSTM) models for both choice problems are estimated (i) for all participants, (ii) for participants exposed to video ads, and (iii) for participants exposed to poster ads. To establish baselines for evaluating the accuracy of these models, we calculate sample averages in the training split and use them to predict choices in the test split.

For the estimations, we excluded 8 participants from the sample due to missing responses in their survey answers. The final sample comprised 95 participants, 48 of whom were exposed to video ads and 47 to poster ads. Table 5 summarizes the participants' characteristics for the complete sample and each exposure type subsample.

4.2. Interpretability

Table 6 presents the logistic regression results of binomial choices on SR, and ME emotion variables. All specifications include ad and product characteristics measurements as controls. The accuracy value reported at the bottom of Table 6 is the proportion of binomial choices where the model's prediction matched the actual observed choice, i.e., the number of correctly predicted choices divided by the total number of observations. Column (1) uses only ad and product characteristics and serves as the starting point for our regression analysis. Columns (2) and (3) compare SR and ME emotions using the complete sample. Columns (4) and (5) compare SR and ME emotions for the participants exposed to video ads, while (6) and (7) for the participants exposed to poster ads.

Table 5
Summary Statistics of Participant Characteristics.

Exposure Type	Nobs	Mean Age	Age SD	Female %	Male %	Other %
Picture	47	30.23	9.45	0.26	0.68	0.06
Video	48	31.92	10.12	0.29	0.71	0
All	95	31.08	9.78	0.27	0.69	0.03

The ad-and-product-characteristics-only case of column (1) has 79% accuracy (percentage of correct predictions in total predictions) in the complete sample. The accuracy in the video sample is 3 percentage points (pp) greater when adding either SR or ME emotions (columns (4), (5) vs. column (1)). This increase is more modest than the improvements observed when incorporating more obtrusive EEG measurements; for example, Hakim et al. (2021) reported a 4.07 pp increase over SR using similar ML methods. Accuracy worsens by 1 pp in the poster sample when SR emotions are used and remains constant with ME emotions (columns (6), (7) vs. column (1)).

Furthermore, we compare the added accuracy contributions of ME and SR emotions over only using ad and product characteristics. Accuracy improves by 1 pp when SR emotions are replaced by ME emotions in the poster ad sample (column (7) vs. column (6)). In the complete and video ad samples, SR and ME emotions yield similar accuracies. The logit results constitute the lower bounds of the ameliorations we observe when SR emotions are replaced by ME emotions for the complete sample and the video ad sample. The subsequent methods we present are more flexible and lead to greater additional accuracy when using ME emotions.

Nonetheless, the logistic regression model has the greatest degree of interpretability among the methods we use and helps develop intuition concerning the relationship between choices and emotions. For instance, column (4) shows that the coefficient of SR disgust is statistically significant at a 1% level, and both columns (3) and (7) show that the coefficients of ME surprise are statistically significant at a 1% level. Since these variables represent differences in emotions triggered by the ads of the two compared products, their signs have natural interpretations. The more surprised (or disgusted) participants appear/report while watching the product *Right* compared to *Left*'s ad, the more (less) likely it

Table 6
Logit with SR and ME Emotions per Exposure Type.

	Dependent variable: Binomial Choice						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Neutral		0.281	3.133**	1.049**	−0.717	−0.009	8.311***
		(0.300)	(1.257)	(0.510)	(1.895)	(0.419)	(2.143)
Happy		0.919**	1.059	0.911	−1.562	1.425**	5.312
		(0.404)	(3.326)	(0.687)	(5.689)	(0.602)	(5.189)
Sad		0.079	3.214	0.996	−1.700	−0.231	7.282
		(0.648)	(3.293)	(1.058)	(5.040)	(0.939)	(5.435)
Angry		−0.987	1.660	−1.365	−2.659	0.437	4.516
		(0.754)	(3.375)	(1.183)	(5.107)	(1.269)	(6.136)
Surprised		0.230	7.483***	0.091	0.083	0.297	13.226***
		(0.297)	(2.368)	(0.456)	(3.832)	(0.463)	(3.567)
Scared		−0.476	−9.910*	1.133	−21.617*	−2.850**	−7.343
		(0.802)	(5.384)	(1.195)	(12.684)	(1.203)	(6.916)
Disgusted		0.082	−0.010	−3.762***	−3.516	2.823***	1.371
		(0.658)	(3.091)	(1.145)	(4.793)	(0.911)	(5.197)
Contempt		0.635	1.008	2.119***	2.380	−0.673	1.655
		(0.440)	(1.475)	(0.686)	(2.279)	(0.677)	(2.098)
Valence		0.800	1.968	1.128	−3.732	0.740	3.030
		(0.512)	(6.245)	(0.885)	(10.016)	(0.727)	(10.312)
Arousal		−0.592	−0.603	−0.594	2.225	−0.882	−1.992
		(0.381)	(0.891)	(0.592)	(1.552)	(0.556)	(1.228)
Constant	−0.215***	−0.210***	−0.207***	−0.255**	−0.230**	−0.188*	−0.216**
	(0.071)	(0.072)	(0.072)	(0.111)	(0.108)	(0.101)	(0.103)
Emotions	None	SR	ME	SR	ME	SR	ME
Exposure	All	All	All	Video	Video	Poster	Poster
Accuracy	0.79	0.79	0.79	0.82	0.82	0.78	0.79
Observations	1,425	1,425	1,425	675	675	690	690
Akaike Inf. Crit.	1,302.120	1,302.617	1,296.775	565.797	591.585	659.727	626.038

Note: All estimations include ad and product characteristics as controls (described in Section 2.1). Columns (2), (4), and (6) use SR emotions. Columns (3), (5), and (7) use ME emotions. Columns (1), (2), and (3) employ the complete sample. Columns (4) and (5) use only participants exposed to video ads. Columns (6) and (7) use only participants exposed to poster ads. The parentheses contain the standard errors. Statistical significance for $p < 0.1$ is denoted with a single star, $p < 0.05$ with two stars, and $p < 0.01$ with three stars. The accuracy is the percentage of choices correctly predicted by the model.

becomes they choose *Right*.

Statistically significant positive effects are found for neutrality (columns 3, 4, and 7), happiness (2 and 6), surprise (3 and 7), and contempt (4). The positive effect of happiness is intuitive. However, the effect of neutrality is more difficult to interpret, as its meaning is highly contextual—unlike emotions such as happiness. For example, the interpretation of neutrality can vary depending on cultural background (Mesquita & Markus, 2004). Intuitive negative effects are estimated for fear (3, 5, and 6). SR Disgust shows an intuitive negative effect in column (4) for the video ad sample, but interestingly, this reverses for the poster ad sample in column (6). The positive coefficient for contempt (column 4) is unexpected. Since contempt is a negative emotion, it is reasonable to expect a negative coefficient, i.e., the more contempt one feels toward the *Right* ad compared to the *Left* ad, the less likely they are to choose *Right*. Notably, both non-intuitive coefficient estimates in Table 6 arise when using SR emotion variables.

4.3. Variable contributions to prediction accuracy

While the logit model of Section 3.2 can be used to interpret the contributions of individual variables on prediction accuracy, it is less informative about how they relate to each other. In this section, we reexamine the contributions of ME, SR, ad, and product characteristics variables, focusing on how influential they are relative to each other when predicting consumer choices. Section 3.3.1 introduces the methodology we employ to assess the relative influence of variable contributions, and Section 3.3.2 presents the results and examines underlying processes that can rationalize the observed differences in predictive power between ME and SR data.

4.3.1. Relative influence

Our analysis quantifies the explanatory contribution of each variable by its relative influence on decision trees representing preferences. Relative influence is calculated as the decrease in the Gini index that introducing each variable in a decision tree induces on average. The Gini index gives a measure of dispersion between the prediction classes. It is given by the formula $G_k = P_{k,l}(1 - P_{k,l}) + P_{k,r}(1 - P_{k,r})$, where $k \in \{l, r\}$ is a prediction class, and $P_{k,l}$ ($P_{k,r}$) is the proportion of *Left* (*Right*) observations classified as k. Small Gini indices indicate that prediction dispersion is low and that each prediction class predominantly contains observations from a single (observed) class.

We present results from decision tree boosting (Greenwell et al., 2022; Liaw & Wiener, 2002). Additional analyses using Lasso, Ridge, and exhaustive subset selection (Friedman et al., 2010) yield similar results and are presented in Web Appendix C for brevity.

4.3.2. Results and insights

Fig. 3 shows the relative influence of each variable using boosting for the average (mean) and extreme (max) time-aggregated ME data. The six most influential variables are product characteristics. All emotion variables measured by ME are more influential than the SR emotion variables. This result is robust when looking at participants exposed to only video or poster ads (see Web Appendix B). SR emotion measurements have only minimal average influence on binary preferences in the estimated decision. Specifically, the average Gini index reduction is less than 2% when SR variables are included in a decision tree.

Fig. 3 corroborates the additional predictive capacity of ME measurements. ME measurements can offer more extensive representations of SR emotions that are already of interest in conventional marketing surveys. However, the results of Fig. 3 do not explain why this effect is in place. What processes can rationalize our results? We put forward two non-mutually exclusive explanations.

First, FEA technologies can capture participants' subconscious facial cues. By definition, participants are not aware of their subconscious

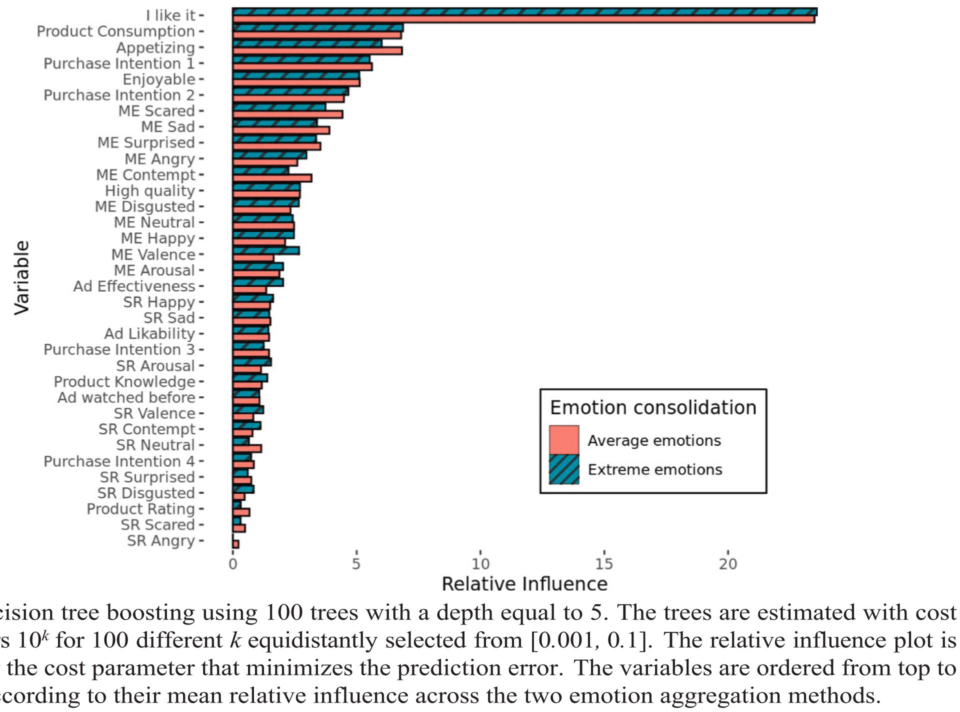


Fig. 3. Decision Tree Boosting on Ad, Product, and Emotion Variables. *Note*: Decision tree boosting using 100 trees with a depth equal to 5. The trees are estimated with cost parameters 10^k for 100 different k equidistantly selected from [0.001, 0.1]. The relative influence plot is drawn for the cost parameter that minimizes the prediction error. The variables are ordered from top to bottom according to their mean relative influence across the two emotion aggregation methods.

processes and, therefore, cannot self-report them. Second, ME measurements are collected during ad exposure, while SR measurements are collected with a time lag. Subtle questions concerning their emotional state during the ads' duration have more ambiguous interpretations than purchase intention questions. Some participants might only focus on the extreme emotions they experienced; others might try to calculate their average emotional state mentally, and others might not perfectly recall their emotions during the ads. Thus, due to self-reporting differences, SR emotion measurements can be less consistent than FEA.

4.4. Classification accuracy

This section quantifies how much ME measurements can improve survey-based choice predictions. We shift our focus from interpretability to more flexible ML methods with greater classification and predictive capabilities. Specifically, we use LDA as a multiproduct classifier for the multinomial choice dataset, and SVM as a binary classifier for the binomial choice dataset (Meyer et al. (2023) and Venables and Ripley (2002); see also Hakim et al. (2021) for similar methodologies with EEG data).

Table 7 summarizes the classification accuracies of LDA and SVM estimations. Columns (1) to (5) present accuracy percentages for estimations using ad and product; SR; ME; ad, product and SR; and ad, product and ME variables, respectively. The last two columns show the differences in accuracy between columns (3) and (2) and columns (5) and (4). The results are presented for the complete sample and the subsamples of participants exposed to video and poster ads.

4.4.1. LDA of multinomial choices

Fig. 4 summarizes the results of LDA estimations for the multinomial choice problem with six product classes. The first two rows display LDA results for the complete sample, the two subsequent rows for participants exposed to video ads, and the last two for participants exposed to poster ads. ME emotions are aggregated as averages over the timeframe of the exposure. The parenthesized percentages in each title give the accuracy of the fit (correct classifications over the number of observations). Cross (red) markers give the product share choices observed in the data. Circle (cyan) markers give the share of observations classified in a product class by LDA.

The method reaches 39.03%, 42.92%, and 45.17% in-sample accuracies in the complete, video, and poster ad participants, respectively, when only ad and product variables are used. Using exclusively SR emotion variables results in lower accuracies for all exposure groups (31.68%, 28.02%, and 38.13%). Accuracy also worsens compared to the ad-and-product-only case when using only ME emotion variables for the video sample (38.47%), but improves for the complete and poster

Table 7
Summary of LDA and SVM Classification Accuracies.

Sample	Ad & Pr.	SR	ME	Ad, Pr. & SR	Ad, Pr. & ME		
	(1)	(2)	(3)	(4)	(5)	(3) − (2)	(5) − (4)
LDA (in-sample)							
All	39.03%	31.68%	39.21%	42.18%	46.23%	7.53 pp	4.05 pp
Video	42.92%	28.02%	38.47%	49.15%	53.64%	10.45 pp	4.49 pp
Poster	45.17%	38.13%	54.93%	57.69%	63.10%	16.80 pp	5.41 pp
SVM (out-of-sample/best validation fold)							
All		64.53%	66.09%	78.29%	86.23%	1.56 pp	7.94 pp
Video		65.22%	65.65%	81.07%	95.96%	0.43 pp	14.89 pp
Poster		75.12%	60.50%	78.82%	79.53%	−14.63 pp	0.71 pp

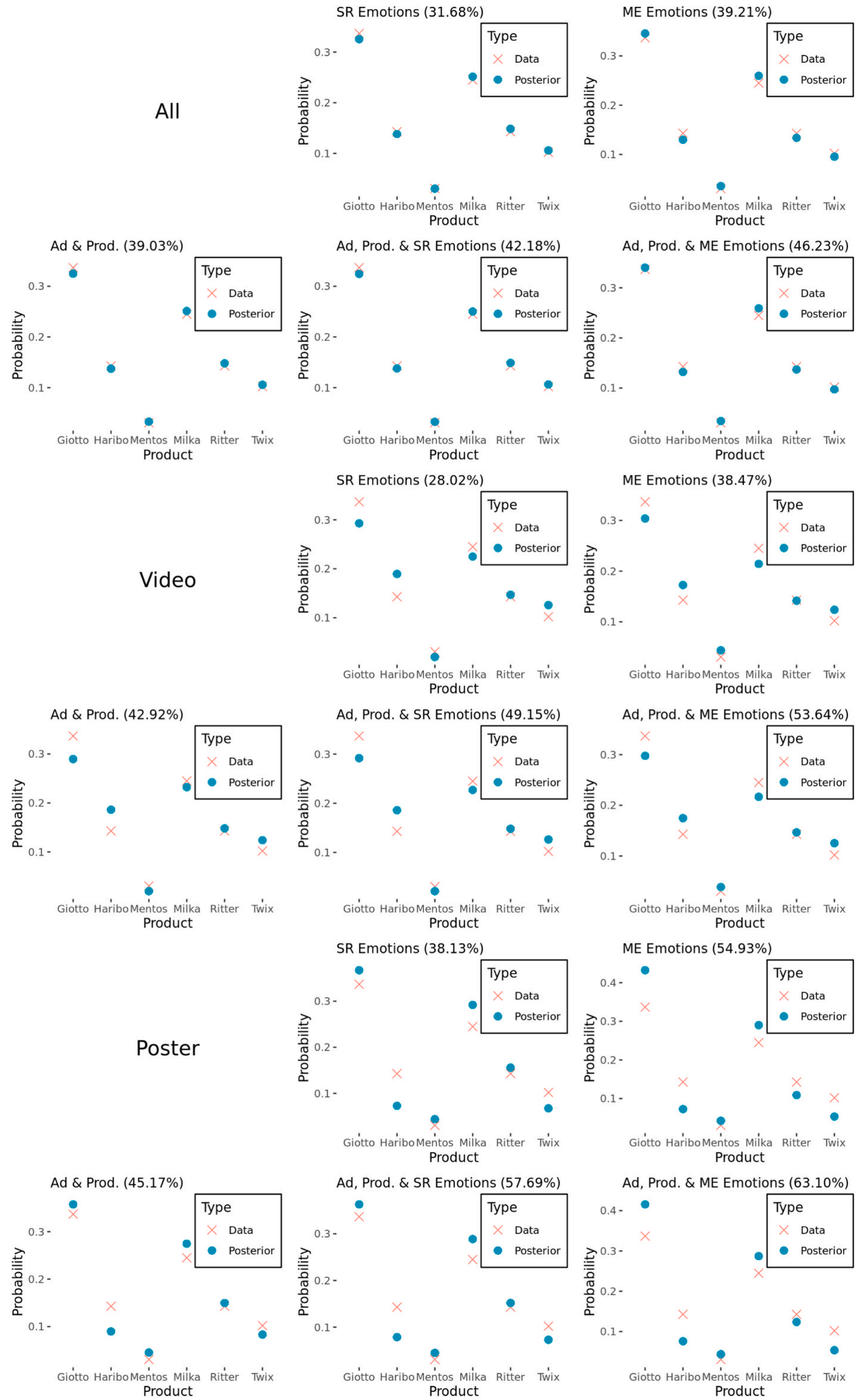


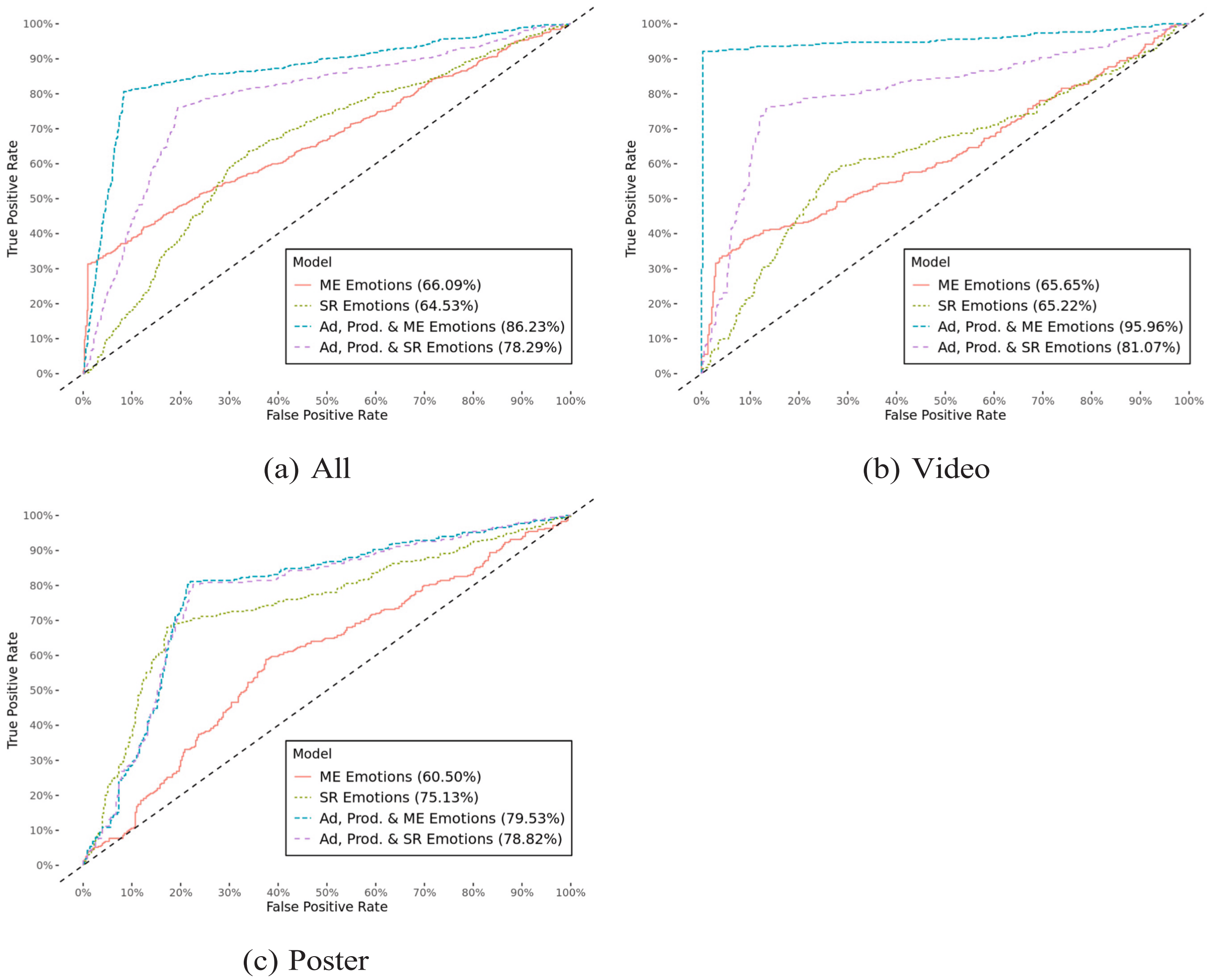
Fig. 4. Multinomial Choice LDA Estimations.

samples (39.21% and 54.93%). When comparing SR-only and ME-only models, accuracy improves by 7.53 pp, 10.45 pp, and 15.80 pp in the complete, video, and poster ad samples, respectively, when ME emotions are used.

The results are similar when emotion variables are combined with ad and product variables. In the complete sample, predictions using ad and product variables and SR emotions have an accuracy of 42.18%. When SR emotions are substituted by ME, accuracy reaches 46.23%, representing a 4.05 pp improvement. For the participants exposed to videos, using ME leads to 4.49 pp more accuracy than using SR. Finally, for the poster ad participants, the improvement from switching to ME emotions is 5.41 pp.

4.4.2. SVM of binomial choices

Classification accuracies across exposure types exhibit similar patterns when applying SVM with a polynomial kernel to the binomial choice problem. Fig. 5 visualizes the out-of-sample classification accuracies per exposure group. The solid (red) lines are obtained by models



Note: Lines curved towards the upper-left corner represent models with better classification properties. The models are obtained as best out-of-sample fits from 15-fold cross-validation over the Cartesian product $C \times D$ of $C = \{0.001, 0.01, 0.1, 1, 10\}$ margin violation cost parameters and $D = \{1, 2, 3, 4, 5\}$ polynomial degrees. The best models' cost parameters are 0.1 (SR), 10 (ME), 0.1 (SSR), 1 (SME) for the complete sample; 1 (SR), 10 (ME), 0.1 (SSR), 10 (SME) for the video; and 1 (SR), 10 (ME), 1 (SSR), 10 (SME) for the poster ad participants. The best models' polynomial degrees are 1 (SR), 5 (ME), 1 (SSR), 3 (SME) for the complete sample; 1 (SR), 3 (ME), 1 (SSR), 3 (SME) for the video; and 3 (SR), 1 (ME), 1 (SSR), 1 (SME) for the poster ad participants. The percentages in parentheses next to the model names in the legend give the classification accuracies of each model fit.

Fig. 5. Receiver Operating Characteristic Curves of SVM for Different Exposure Types. *Note*: Lines curved towards the upper-left corner represent models with better classification properties. The models are obtained as best out-of-sample fits from 15-fold cross-validation over the Cartesian product $C \times D$ of $C = \{0.001, 0.01, 0.1, 1, 10\}$ margin violation cost parameters and $D = \{1, 2, 3, 4, 5\}$ polynomial degrees. The best models' cost parameters are 0.1 (SR), 10 (ME), 0.1 (SSR), 1 (SME) for the complete sample; 1 (SR), 10 (ME), 0.1 (SSR), 10 (SME) for the video; and 1 (SR), 10 (ME), 1 (SSR), 10 (SME) for the poster ad participants. The best models' polynomial degrees are 1 (SR), 5 (ME), 1 (SSR), 3 (SME) for the complete sample; 1 (SR), 3 (ME), 1 (SSR), 3 (SME) for the video; and 3 (SR), 1 (ME), 1 (SSR), 1 (SME) for the poster ad participants. The percentages in parentheses next to the model names in the legend give the classification accuracies of each model fit.

using only ME variables, the dotted (green) lines using only SR variables, the dash-dotted (cyan) lines using ad, product and ME variables, and the dashed (purple) lines using ad, product and SR variables. The SVM cost and polynomial degree parameters are determined through 15-fold cross-validation, using the accuracies of the models on the test splits as the selection criterion.

Models that use only ME emotion variables achieve higher accuracy than those using only SR emotion variables in the complete sample (by 1.56 pp) and the video ad sample (by 0.43 pp). In contrast, the SR-only model has a 14.63 pp accuracy advantage in the poster ad subsample. However, models that combine ME, ad and product variables consistently outperform those combining SR, ad and product variables across all exposure types: with improvements of 7.94 pp in the complete sample, 14.89 pp in the video subsample, and 0.71 pp in the poster ad subsample.

4.5. ME dynamics

We examine whether utilizing the dynamics of the ME data can further improve choice prediction accuracy. In contrast to ME data, SR data is atemporal. Thus, we employ two classes of ANN models in our analysis to accommodate for this difference between SR and ME data. Static DL models are used for estimating multinomial and binomial choice models with SR data. Dynamic LSTM models are used for estimating multinomial and binomial choice models with ME data. All models include ad and product characteristics. For the technical details of the model architectures, we refer the reader to Web Appendix D.

Table 8 presents the learning, data, and evaluation parameters used in the ANN model estimations. We use K-fold cross-validation to evaluate the models' performance. Each model is estimated K times, with each fold iteration using a different random split of the data. We construct three sample splits for each iteration. The test split contains 20% of the participants, the validation split contains 32% of the participants, and the training split contains the remaining 48% of the participants. The training split is used to fit the models, the validation split is used to establish train-stopping conditions for avoiding overfitting, and the test split is used for evaluating the fitted models. There is no overlap of participants between the three splits. Thus, the models are evaluated against participant observations that are absent from the training data.

The baseline and multinomial choice models are estimated using $K = 100$ folds. The binomial choice models are estimated using 10 folds due to the scale of their computational requirements.

The baseline choice rules for the evaluation of the ANN models are calculated by averages in the training data splits. That is, for each fold iteration, we calculate (1) the shares of product choices across participants in the training split, $p_R = (p_{R,1}, \ldots, p_{R,6})$, (2) the shares of product choices across participants in the test split, $p_T = (p_{T,1}, \ldots, p_{T,6})$, and (3) the inner product $E = p_R \cdot p_T$. The value of E gives the expected prediction accuracy when expectations are formed according to the training shares p_R. We use E to establish a minimum accuracy baseline above which it is meaningful to use the ANN models for choice predictions.

Fig. 6 visualizes the results of the K-fold estimations. Fig. 6a displays the empirical densities and averages of the test accuracies obtained from the cross-validation folds of the multinomial choice models. Fig. 6b shows the densities and averages from the folds of the binomial choice models. The prediction accuracy improvements over the baseline are more pronounced in the binomial choice models irrespective of the emotion measurements. For both choice models, using ME emotions leads to greater accuracy improvements for the complete and video ad participant samples. ME emotions also give greater improvements for the poster ad sample of the binomial choice model, but not for the multinomial choice model. Average accuracies and differences between SR, ME, and baseline estimations are summarized in Table 9. Parenthesized values in the table give the p-values of the paired t-tests comparing the test accuracies of the models.

Table 8
ANN Estimation Parameters.

	Baseline		Multinomial		Binomial	
	SR	ME	SR	ME	SR	ME
Learning						
Batch Size	–	–	256	256	256	256
LSTM Window	–	–	–	25	–	25
Learning Rate	–	–	10^{-3}	10^{-3}	10^{-3}	10^{-3}
Decay Rate	–	–	10^{-2}	10^{-2}	10^{-2}	10^{-2}
Early Stopping	–	–	10^{2}	10^{2}	10^{2}	10^{2}
Max Epochs	–	–	10^{5}	10^{5}	10^{5}	10^{5}
Data						
Train Split	48%	48%	48%	48%	48%	48%
Validation Split	32%	32%	32%	32%	32%	32%
Test Split	20%	20%	20%	20%	20%	20%
Evaluation						
K-Fold	100	100	100	100	10	10

The average baseline accuracies from Fig. 6a are approximately 20.93% for the full sample, 23.57% for the poster ad subsample, and 19.15% for the video ad subsample. For the complete sample, the highest average accuracy is achieved using ME emotions (30.19%), representing a 9.26 pp improvement over the baseline and a 2.45 pp improvement over SR-based predictions. Similarly, ME-based predictions are most accurate for participants exposed to video ads (22.96%), with a 3.81 pp increase over the baseline and a 3.63 pp improvement with a low p-value of 0.0283 over SR-based predictions. In contrast, SR-based predictions perform better for participants exposed to poster ads, reaching 25.78% accuracy, that is, 2.48 pp higher than ME-based predictions and 2.21 pp above the baseline.

In Fig. 6b, ME-based predictions achieve the highest accuracy across all exposure types. However, the differences between ME- and SR-based predictions have high p-values. Oppositely, both ME- and SR-based models significantly outperform baseline predictions. The average baseline accuracies are approximately 50.65% for the complete sample, 50.87% for the poster ad subsample, and 50.46% for the video ad subsample. For the complete sample, the highest accuracy is achieved using ME emotions (74.28%), which is 23.63 pp above the baseline and 0.32 pp higher than SR-based predictions. For poster ads, ME-based predictions again yield the highest accuracy (71.70%), improving on the baseline by 20.83 pp and outperforming SR by 0.37 pp. Finally, for video ads, ME-based predictions reach 74.31% accuracy, 23.85 pp above the baseline and 0.44 pp higher than SR-based predictions.

5. Discussion

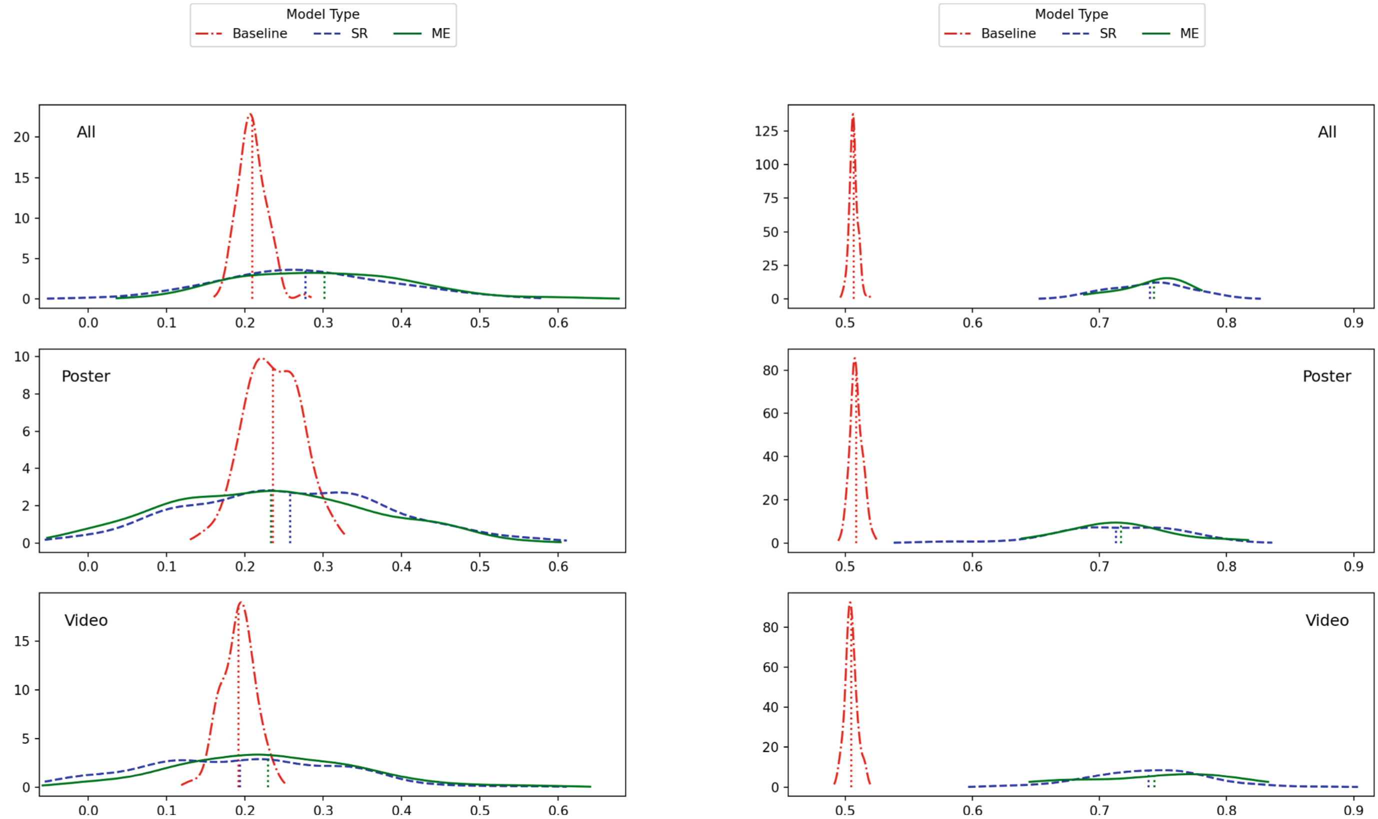
This study employed ML and ANN methodologies to delve into consumer choice predictions by utilizing FEA. Our approach demonstrated that augmenting traditional marketing surveys with automatic FEA can circumvent potential biases and inherent limitations of SR emotion data. This fusion of techniques offers a novel pathway for eliciting consumer preferences in marketing research. Moreover, integrating face-reading technology into surveys enables access to real-time emotion measurements during ad exposure. Our prediction models, grounded in ME, present a cost-effective and scalable alternative to more resource-intensive measurements like EEG and fMRI, making them applicable to a larger participant pool within a condensed timeframe.

5.1. Contributions and limitations

Summarizing the results, we conclude with five major findings. First, we focus on the conceptual insights of the study and, second, on its technical innovations.

On the conceptual side, we find (1) that across both multinomial and binomial scenarios, ME-based emotion data can improve choice prediction accuracy over SR-based data, whether used alone or alongside ad and product characteristics. These improvements hold consistently across a range of ML and ANN models of varying complexity. (2) Our evidence suggests that ME emotions obtained via FEA can provide a more nuanced and accurate depiction of consumer emotions compared to SR measurements. We propose two non-mutually exclusive explanations of this result. First, FEA can capture subconscious facial cues that participants cannot report as they are not fully aware of them. Second, ME offer dynamic emotion measurements collected during ad exposure, while SR are usually retrospective, single data-point measurements that are potentially subject to self-reporting biases. (3) The benefits of employing ME are more pronounced for video ads. Overall, the prediction improvement differences between ME and SR are smaller for poster than for video ads. Video ads, being more dynamic, evoke complex emotional reactions that are harder to capture in a single SR response. MEs offer a richer emotional profile due to their dynamic nature.

On the technical side, (4) incorporating ME data into ANN-based



(a) Multinomial Choice. (b) Binomial Choice.

Fig. 6. ANN Model Test Accuracies across Exposure Types.

Table 9
Summary of ANN Test Accuracies.

Sample	Baseline	SR	ME	ME-SR	SR-Baseline	ME-Baseline
Multinomial						
All	20.93%	27.74%	30.19%	2.45 pp (0.1036)	6.81 pp (0.0000)	9.26 pp (0.0000)
Poster	23.57%	25.78%	23.30%	−2.48 pp (0.1663)	2.21 pp (0.0917)	−0.27 pp (0.8388)
Video	19.15%	19.33%	22.96%	3.63 pp (0.0283)	0.18 pp (0.8821)	3.81 pp (0.001)
Binomial						
All	50.65%	73.96%	74.28%	0.32 pp (0.422)	23.31 pp (0.0000)	23.63 pp (0.0000)
Poster	50.87%	71.33%	71.70%	0.37 pp (0.5573)	20.46 pp (0.0000)	20.83 pp (0.0000)
Video	50.46%	73.87%	74.31%	0.44 pp (0.5316)	23.41 pp (0.0000)	23.85 pp (0.0000)

choice models enhances prediction accuracy compared to baseline choice rules. While both ME and SR contribute positively, ME exhibit greater accuracy improvements. Notably, the improvements over the baseline rules are stronger in binomial choice scenarios. (5) Replacing SR with ME in the proposed ANN framework increases out-of-sample prediction accuracy across all exposure types in the binomial choice setting. For multinomial choices, replacing SR with ME improves accuracy for video ad viewers but not for poster ad viewers. In particular, ME improves accuracy by 3.63 pp over SR for video ad viewers; a statistically significant difference at the 5% level.

Finally, we identify certain aspects that our approach leaves unanswered and that future research could address. First, we focused on confectionery products, which are affordable and appealing across various demographics, as we aimed to examine the predictability of take-out choices of participants. A question left open from this approach, however, is whether the prediction accuracy improvements we find from involving FEA measurements are robust for less affordable products. Second, using confectionery products additionally allowed us to gather measurements from a diverse set of brands. This diversity increases the ecological validity of our results compared to focusing on a single brand but introduces the possibility of overshadowing effects, as participants were exposed to multiple ads in a short time frame. Future research could explore designs that collect SR measurements from fewer ads or immediately after ad exposure to mitigate this effect.

We additionally document three technical aspects that future work could consider. First, our study does not investigate the optimal time window for the LSTM models. The time window is set to 5 s in our approach to ensure it captures at least a substantial range of the exposure durations for the ads in our experiment. However, the optimal time window can differ for different participants and products. Second, the ANN models we use are relatively shallow. Deeper networks have a greater capacity to represent more complex relationships in the data. However, deeper networks are also more prone to the vanishing gradient problem. Future research can investigate network depth's effects on our models' prediction accuracy. Third, our investigation of the accuracy gains from incorporating the time dimension of ME measurements uses only LSTMs. By design, LSTMs are autoregressive models with a fixed time window preceding the current observation. However, consumers exposed to ads may formulate emotional states by processing information from different parts of the ad that are not necessarily contiguous in time. Future research can investigate the choice prediction accuracy using transformer models with attention mechanisms that can capture the non-contiguous nature of emotional states. This may also be impactful in cases where the primary process of forming an emotional state lies in the first impressions from the exposure, such as in the case of poster ads.

5.2. Implications in management and research

These findings carry significant implications for both industry practitioners and researchers. For companies and brands, our approach offers a refined method to inform product and service development, as well as to guide the design of marketing and advertising campaigns, enhancing their effectiveness in eliciting consumer preferences. The advantages of this approach can be especially pronounced in private, sensitive, or potentially embarrassing product or service domains, where consumers may hesitate to openly communicate their preferences (e.g., hygiene products or political affiliations).

Integrating FEA into ANN models allows for dynamic testing and design-adaptation of ads. Instead of relying solely on traditional A/B testing, our results suggest that marketers can use real-time ME data to more accurately predict how alternative ad designs affect consumer choices. In turn, this can enable marketers to adjust ad design more flexibly, achieving much faster iterations in design-test-learn cycles. This approach is particularly beneficial for video ads, where MEs provide the ability to track emotions dynamically across ad exposure and lead to more accurate predictions of consumer choices. Thus, our approach can help brand ad creatives to identify winning ad variations more quickly and optimize their messaging.

The improvements in prediction accuracy can be practically impactful in real marketing applications. Gains in accuracy can translate into financial and strategic advantages for companies (Bayer et al., 2020; Brandt, 2016; Garaus et al., 2021). For instance, in large-scale advertising campaigns, an improvement in choice prediction accuracy can lead to better-targeted ads and, eventually, to a larger share of consumers being correctly identified as likely buyers. This, in turn, can lead to higher conversion rates and more efficient allocation of marketing budgets.

For researchers, the integration of automatic FEA can streamline data collection by reducing reliance on extensive questionnaires. Since facial expressions can be monitored during stimulus exposure, there is less need for additional emotion-related survey questions, thereby conserving responders and interviewers' time. Simultaneously, FEA integration enables interviewers to receive immediate feedback and predictions about subject preferences and dynamically adapt the interviews accordingly. However, the potential of FEA extends beyond improving prediction accuracy. As customer journeys become increasingly complex, traditional market research tools may prove inadequate to leverage consumer reactions throughout these processes fully. Thus, it is essential to explore further the dynamic nature of FEA and its potential to enhance market research practices longitudinally.

5.3. Social and ethical implications

As the use of automated facial expression analysis (FEA) in marketing research grows, it is important to consider both conceptual and ethical aspects of this technology. FEA differs fundamentally from facial recognition technology (FRT): while FRT identifies or verifies individuals, FEA interprets facial movements to infer emotional expressions. Importantly, FEA does not directly measure consumers' emotional states; these inferences can be influenced by cultural norms, display rules, and algorithmic biases, which may lead to misinterpretation or oversimplification of emotional responses (Barrett, Adolphs, Marsella, Martinez, & Pollak, 2019). Recognizing these limitations is essential for responsibly applying FEA in commercial contexts and provides a foundation for the broader discussion of ethical and privacy considerations that follow.

Extracting information on preferences through ANNs and FEA entails a shift in bargaining power from the information provider to the recipient. In many instances, such shifts raise ethical and safety concerns. For example, eliciting preferences from non-consenting consumers exposed to street ads may entail elements of privacy invasion. Given these considerations, amendments to regulations concerning the use of personal visual data in commercial activities may be warranted.

The US and China, which lead AI technological innovations at the time of writing this article, lack an overarching AI regulation framework. Instead, both the US and China shift emphasis toward removing innovation barriers and adopt more market-regulating approaches for AI use.

The EU AI Act entered into force on August 1, 2024, and will gradually become fully applicable on August 2, 2026. The Act establishes a comprehensive legal framework for AI technologies, aiming to balance innovation with ethical considerations. Among other provisions, the Act mandates that tech companies impose restrictions on practices like indiscriminate image scraping for facial recognition databases. Nonetheless, concerning the use of facial recognition technologies, the Act primarily focuses on law enforcement and surveillance aspects. Less focus is placed on market and consumer applications using FEA and ANN prediction algorithms.

In stark contrast, safety aspects and ethical concerns of consumer AI applications are emphasized by the international AI engineering community. For example, ISO/IEC 42001 (2023) sets standards for AI system management relating to safety, transparency, and fairness. Such standards are relevant for potential commercial applications using FEA data. Further, ISO/IEC TR 24368 (2022) provides guidelines for the ethical use of AI systems and ISO/IEC TR 24027 (2021) for mitigating biases in AI-aided decision-making systems. Seminal work on AI biases originally focused on gender (Bolukbasi et al., 2016; Buolamwini & Gebru, 2018), but active research and the recommendations of ISO/IEC TR 24027 (2021) have a wider view of biases, going beyond gender discrimination.

The models we propose in Section 3.5 can support marketing decision-making in ad design processes. In addition, they have the potential for unconscious preference elicitation from ME data. These characteristics entail ethical implications even for firms aiming to be legally compliant, as they can inherently shift bargaining power away from consumers. Further, they raise concerns about the potential for exploitation of preference predictions by malicious actors using non-consensually collected ME data. These concerns are exacerbated both by the current lack of relevant international regulation and by the low entry barriers for using such models, as they rely solely on hardware with video recording capabilities, which is both cheap and widely available. Therefore, we find it necessary to advocate for careful consideration of the ethical implications when using FEA in commercial applications.

CRediT authorship contribution statement

Pantelis Karapanagiotis: Writing – review & editing, Writing – original draft, Project administration, Methodology, Formal analysis, Data curation, Conceptualization. **Franziska Krause:** Writing – review & editing, Writing – original draft, Software, Resources, Project administration, Data curation, Conceptualization. **Janina Krick:** Writing – review & editing, Writing – original draft, Software, Project administration, Methodology, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jbusres.2026.116050.

Data availability

The links to the data and code are included in the manuscript as well as in the Letter to the Editor.

References

Ariely, D., & Berns, G. S. (2010). Neuromarketing: The hope and hype of neuroimaging in business. *Nature Reviews Neuroscience, 11*(4), 284–292. https://doi.org/10.1038/nrn2795

Bagozzi, R. P., Gopinath, M., & Nyer, P. U. (1999). The Role of Emotions in Marketing. *Journal of the Academy of Marketing Science, 27*(2), 184–206. https://doi.org/10.1177/0092070399272005

Barrett, L. F., Adolphs, R., Marsella, S., Martinez, A. M., & Pollak, S. D. (2019). Emotional expressions reconsidered: Challenges to inferring emotion from human facial movements. *Psychological Science in the Public Interest, 20*(1), 1–68.

Bartlett, M. S., Littlewort, G. C., Frank, M. G., Lainscsek, C., Fasel, I. R., & Movellan, J. R. (2006). Automatic Recognition of Facial Actions in Spontaneous Expressions. *Journal of Multimedia, 1*(6), 22–35. https://doi.org/10.4304/jmm.1.6.22-35

Bayer, E., Srinivasan, S., Riedl, E. J., & Skiera, B. (2020). The impact of online display advertising and paid search advertising relative to offline advertising on firm performance and firm value. *International Journal of Research in Marketing, 37*(4), 789–804. https://doi.org/10.1016/j.ijresmar.2020.02.002

Bazzani, A., Ravaioli, S., Trieste, L., Faraguna, U., & Turchetti, G. (2020). Is EEG Suitable for Marketing Research? *A Systematic Review. Frontiers in neuroscience, 14*, Article 594566. https://doi.org/10.3389/fnins.2020.594566

Blasco-Arcas, L., Lee, H.-H. M., Kastanakis, M. N., Alcañiz, M., & Reyes-Menendez, A. (2022). The role of consumer data in marketing: A research agenda. *Journal of Business Research*, *146*, 436–452. Doi: 10.1016/j.jbusres.2022.03.054.

Bolukbasi, T., Chang, K.-W., Zou, J., Saligrama, V., & Kalai, A. (2016). Man is to computer programmer as woman is to homemaker? Debiasing word embeddings. *Proceedings of the 30th International Conference on Neural Information Processing Systems*, 4356–4364. Doi: arXiv:1607.06520.

Brandt, D. (2016). *What's Next: Emotions Give a Lift to Advertising* (tech. rep.). Nielsen. Retrieved August 3, 2025, from https://www.nielsen.com/insights/2016/whats-next-emotions-give-a-lift-to-advertising/.

Buchanan, B., & Henderson, P. W. (1992). Assessing the Bias of Preference, Detection, and Identification measures of Discrimination Ability in Product Design. *Marketing Science, 11*(1), 64–75. https://doi.org/10.1287/mksc.11.1.64

Buolamwini, J., & Gebru, T. (2018, February). Gender shades: Intersectional accuracy disparities in commercial gender classification. In S. A. Friedler & C. Wilson (Eds.), *Proceedings of the 1st conference on fairness, accountability and transparency* (pp. 77–91, Vol. 81). PMLR. https : / / proceedings . mlr . press / v81 / buolamwini18a. html.

Camerer, C., & Yoon, C. (2015). Introduction to the *Journal of Marketing Research* Special Issue on Neuroscience and Marketing. *Journal of Marketing Research*, *52* (4), 423–426. Doi: 10.1509/0022-2437-52.4.423.

Cohn, J. F., & De La Torre, F. (2015, January). Automated Face Analysis for Affective Computing. In R. Calvo, S. D'Mello, J. Gratch, & A. Kappas (Eds.), *The Oxford Handbook of Affective Computing*. Oxford University Press. Doi: 10. 1093/oxfordhb/9780199942237.013.020.

Darwin, C. (1965/1872). *The expression of the emotions in man and animals*. University of Chicago Press. Carrier: Lorenz K.

De Wijk, R., Kaneko, D., Dijksterhuis, G., Van Bergen, G., Vingerhoeds, M., Visalli, M., & Zandstra, E. (2022). A preliminary investigation on the effect of immersive consumption contexts on food-evoked emotions using facial expressions and subjective ratings. *Food Quality and Preference, 99*, Article 104572. https://doi.org/10.1016/j.foodqual.2022.104572

Den Uyl, M., & Van Kuilenburg, H. (2005). The FaceReader: Online facial expression recognition. *Proceedings of Measuring Behavior*, *30*.

Derbaix, C. (1995). The Impact of Affective Reactions on attitudes toward the Advertisement and the brand: A step toward Ecological Validity. *Journal of Marketing Research, 32*(4), 470–479. https://doi.org/10.1177/002224379503200409

Derbaix, C., & Bree, J. (1997). The impact of children's affective reactions elicited by commercials on attitudes toward the advertisement and the brand. *International Journal of Research in Marketing, 14*(3), 207–229. https://doi.org/10.1016/S0167-8116(97)00003-7

Dimberg, U., Thunberg, M., & Elmehed, K. (2000). Unconscious Facial Reactions to Emotional Facial Expressions. *Psychological Science*, *11* (1), 86–89. https://doi. org/10.1111/1467-9280.00221.

Dixon, D., & Mikolon, S. (2021). Cents of self: How and when self-signals influence consumer value derived from choices of green products. *International Journal of Research in Marketing, 38*(2), 365–386. https://doi.org/10.1016/j.ijresmar.2020.08.002

Donovan, R. J., Rossiter, J. R., Marcoolyn, G., & Nesdale, A. (1994). Store atmosphere and purchasing behavior. *Journal of Retailing, 70*(3), 283–294. https://doi.org/10.1016/0022-4359(94)90037-X

Du, R. Y., Joo, M., & Wilbur, K. C. (2019). Advertising and brand attitudes: Evidence from 575 brands over five years. *Quantitative Marketing and Economics, 17*(3), 257–323. https://doi.org/10.1007/s11129-018-9204-6

Dupré, D., Krumhuber, E. G., Küster, D., & McKeown, G. J. (2020). A performance comparison of eight commercially available automatic classifiers for facial affect recognition (S. D'Mello, Ed.). *PLOS ONE*, *15* (4), e0231968. Doi: 10. 1371/journal. pone.0231968.

Ekman, P. (1992). An argument for basic emotions. *Cognition and Emotion, 6*(3–4), 169–200. https://doi.org/10.1080/02699939208411068

Ekman, P., & Friesen, W. V. (1971). Constants across cultures in the face and emotion. *Journal of Personality and Social Psychology, 17*(2), 124–129. https://doi.org/10.1037/h0030377

Ekman, P., & Friesen, W. V. (1978). Facial Action Coding. *System*. https://doi.org/10.1037/t27734-000

Friedman, J., Hastie, T., & Tibshirani, R. (2010). Regularization Paths for Generalized Linear Models via Coordinate Descent. *Journal of Statistical Software, 33*(1). https://doi.org/10.18637/jss.v033.i01

Garaus, M., Wagner, U., & Rainer, R. C. (2021). Emotional targeting using digital signage systems and facial recognition at the point-of-sale. *Journal of Business Research, 131*, 747–762. https://doi.org/10.1016/j.jbusres.2020.10.065

Gasper, K., Hu, D., & Haynes, E. (2024). Feeling more neutral? Evaluative conditioning can increase neutral affective reactions. *Cognition and Emotion*, 1–19. https://doi.org/10.1080/02699931.2024.2372385

Ghiassaleh, A., Kocher, B., & Czellar, S. (2020). Best seller!? Unintended negative consequences of popularity signs on consumer choice behavior. *International Journal of Research in Marketing, 37*(4), 805–820. https://doi.org/10.1016/j.ijresmar.2020.04.003

González-Rodríguez, M., Díaz-Fernández, M., & Pacheco Gómez, C. (2020). Facial-expression recognition: An emergent approach to the measurement of tourist satisfaction through emotions. *Telematics and Informatics*, *51*, 101404. Doi: 10. 1016/j.tele.2020.101404.

Greenwell, B., Boehmke, B., & Cunningham, J. (2022). *Gbm: Generalized boosted regression models. Manual*. https://CRAN.R-project.org/package=gbm.

Griffin, A., & Hauser, J. R. (1993). The Voice of the Customer. *Marketing Science, 12*(1), 1–27. https://doi.org/10.1287/mksc.12.1.1

Hakim, A., Klorfeld, S., Sela, T., Friedman, D., Shabat-Simon, M., & Levy, D. J. (2021). Machines learn neuromarketing: Improving preference prediction from self-reports using multiple EEG measures and machine learning. *International Journal of Research in Marketing, 38*(3), 770–791. https://doi.org/10.1016/j.ijresmar.2020.10.005

Haley, R. I., & Baldinger, A. L. (2000). The ARF Copy Research Validity Project. *Journal of Advertising Research, 40*(6), 114–140. https://doi.org/10.2501/jar-40-6-114-135

Hazlett, R. L., & Hazlett, S. Y. (1999). Emotional response to television commercials: Facial EMG vs. self-report. *Journal of Advertising Research, 39*(2), 7–23.

Höfling, T. T. A., & Alpers, G. W. (2023). Automatic facial coding predicts self-report of emotion, advertisement and brand effects elicited by video commercials. *Frontiers in Neuroscience*, *17*, 1125983. Doi: 10.3389/fnins.2023.1125983.

Holbrook, M., & Batra, R. (1988). Toward a standardized emotional profile (SEP) useful in measuring responses to the nonverbal components of advertising. In D. Hecker S. & Stewart (Ed.), *Nonverbal communication in advertising* (pp. 95–109). D.C. Heath.

ISO/IEC 42001. (2023, December). *Information technology - Artificial intelligence - Management system* (1st ed., ISO/IEC JTC 1/SC 42 No. Stage 60.60). International Organization for Standardization. Geneva, Switzerland. https://www.iso. org/standard/81230.html.

ISO/IEC TR 24027. (2021, November). *Information technology - Artificial intelligence (AI) - Bias in AI systems and AI aided decision making* (1st ed., ISO/IEC JTC 1/SC 42 No. Stage 60.60). International Organization for Standardization. Geneva, Switzerland. https://www.iso.org/standard/77607.html.

ISO/IEC TR 24368. (2022, August). *Information technology - Artificial intelligence - Overview of ethical and societal concerns* (1st ed., ISO/IEC JTC 1/SC 42 No. Stage 60.60). International Organization for Standardization. Geneva, Switzerland. https://www.iso.org/standard/78507.html.

Johansson, P., Hall, L., Sikström, S., Tärning, B., & Lind, A. (2006). How something can be said about telling more than we can know: On choice blindness and introspection. *Consciousness and Cognition*, *15* (4), 673–692. Doi: 10.1016/j. concog.2006.09.004.

Keller, K. L. (2012). Strategic brand management: Building, measuring, and managing brand equity (4th ed.) Pearson.

Kerrihard, A. L., Khair, M. B., Blumberg, R., Feldman, C. H., & Wunderlich, S. M. (2017). The effects of acclimation to the United States and other demographic factors on responses to salt levels in foods: An examination utilizing face reader technology. *Appetite, 116*, 315–322. https://doi.org/10.1016/j.appet.2017.05.004

Kong, W., Zhao, X., Hu, S., Vecchiato, G., & Babiloni, F. (2013). Electronic evaluation for video commercials by impression index. *Cognitive Neurodynamics, 7*(6), 531–535. https://doi.org/10.1007/s11571-013-9255-z

Kulke, L., Feyerabend, D., & Schacht, A. (2020). A Comparison of the Affectiva iMotions Facial Expression Analysis Software with EMG for Identifying Facial Expressions of Emotion. *Frontiers in Psychology, 11*, 329. https://doi.org/10.3389/fpsyg.2020.00329

Kushwah, S., Dhir, A., & Sagar, M. (2019). Understanding consumer resistance to the consumption of organic food. a study of ethical consumption, purchasing, and choice behaviour. *Food Quality and Preference, 77*, 1–14. https://doi.org/10.1016/j.foodqual.2019.04.003

Lerner, J. S., Li, Y., Valdesolo, P., & Kassam, K. S. (2015). Emotion and Decision making. *Annual Review of Psychology, 66*(1), 799–823. https://doi.org/10.1146/annurev-psych-010213-115043

Lewinski, P., Fransen, M. L., & Tan, E. S. H. (2014). Predicting advertising effectiveness by facial expressions in response to amusing persuasive stimuli. *Journal of*

Neuroscience, Psychology, and Economics, 7(1), 1–14. https://doi.org/10.1037/npe0000012

Li, S., Walters, G., Packer, J., & Scott, N. (2018). A Comparative Analysis of Self- Report and Psychophysiological measures of Emotion in the Context of Tourism advertising. *Journal of Travel Research, 57*(8), 1078–1092. https://doi.org/10.1177/0047287517733555

Liaw, A., & Wiener, M. (2002). Classification and regression by randomForest. *R News, 2* (3), 18–22. https://CRAN.R-project.org/doc/Rnews/.

Loijens, L., & Krips, O. (2025). FaceReader methodology note. Retrieved February 5, 2025, from https://www.noldus.com/download/document/white-paper-facereader-methodology-note.

MacKenzie, S. B., & Podsakoff, P. M. (2012). Common method bias in marketing: Causes, mechanisms, and procedural remedies. *Journal of Retailing, 88*(4), 542–555. https://doi.org/10.1016/j.jretai.2012.08.001

Masip, D., North, M. S., Todorov, A., & Osherson, D. N. (2014). Automated Prediction of Preferences Using Facial Expressions (O. D. Suarez, Ed.). *PLoS ONE*, *9* (2), e87434. Doi: 10.1371/journal.pone.0087434.

McDaniel, S. W., & Kolari, J. W. (1987). Marketing Strategy Implications of the Miles and Snow Strategic Typology. *Journal of Marketing*, *51* (4), 19–30. https://doi. org/10.1177/002224298705100403.

McDuff, D., El Kaliouby, R., Senechal, T., Amr, M., Cohn, J. F., & Picard, R. (2013). Affectiva-MIT Facial Expression Dataset (AM-FED): Naturalistic and Spontaneous Facial Expressions Collected In-the-Wild. *IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2013*, 881–888. https://doi.org/10.1109/CVPRW.2013.130

Mehta, A., Sharma, C., Kanala, M., Thakur, M., Harrison, R., & Torrico, D. D. (2021). Self-Reported Emotions and Facial Expressions on Consumer Acceptability: A Study Using Energy Drinks. *Foods*, *10* (2), 330. https : / / doi . org / 10 . 3390 / foods10020330.

Mesquita, B., & Markus, H. R. (2004). April). Culture and Emotion: Models of Agency as sources of Cultural Variation in Emotion. In A. S. R. Manstead, N. Frijda, & A. Fischer (Eds.), *Feelings and Emotions* ((1st ed.,, pp. 341–358). Cambridge University Press. https://doi.org/10.1017/CBO9780511806582.020.

Meyer, D., Dimitriadou, E., Hornik, K., Weingessel, A., & Leisch, F. (2023). *E1071: Misc functions of the department of statistics, probability theory group (formerly: E1071)*. Manual: TU Wien. https://CRAN.R-project.org/package=e1071.

Neeley, S., & Cronley, M. L. (2004). When research participants DonT tell it like it is: Pinpointing the effects of social desirability bias using self vs. *Indirect-questioning. ACR North American Advances. https : / / api. semanticscholar. org / CorpusID, 147891098*.

Ozretic-Dosen, D., Fuduric, M., & Horvat, S. (2022). Challenging the status quo in marketing research. *Journal of Business Research, 143*, 294–297. https://doi.org/10.1016/j.jbusres.2022.01.077

Pieters, R., & Wedel, M. (2004). Attention Capture and transfer in Advertising: Brand, Pictorial, and Text-size Effects. *Journal of Marketing, 68*(2), 36–50. https://doi.org/10.1509/jmkg.68.2.36.27794

Plassmann, H., Ramsøy, T. Z., & Milosavljevic, M. (2012). Branding the brain: A critical review and outlook. *Journal of Consumer Psychology*, *22* (1), 18–36. https://doi. org/10.1016/j.jcps.2011.11.010.

Poels, K., & Dewitte, S. (2006). How to Capture the Heart? reviewing 20 Years of Emotion Measurement in Advertising. *Journal of Advertising Research, 46*(1), 18–37. https://doi.org/10.2501/S0021849906060041

Ravaja, N., Somervuori, O., & Salminen, M. (2013). Predicting purchase decision: The role of hemispheric asymmetry over the frontal cortex. *Journal of Neuroscience, Psychology, and Economics, 6*(1), 1–13. https://doi.org/10.1037/a0029949

Richins, M. L. (1997). Measuring Emotions in the Consumption Experience. *Journal of Consumer Research, 24*(2), 127–146. https://doi.org/10.1086/209499

Sato, Y., Horaguchi, Y., Vanel, L., & Shioiri, S. (2022). Prediction of image Preferences from Spontaneous Facial Expressions. *Interdisciplinary Information Sciences, 28*(1), 45–53. https://doi.org/10.4036/iis.2022.A.02

Scherer, K. R., & Ellgring, H. (2007). Multimodal expression of emotion: Affect programs or componential appraisal patterns? *Emotion, 7*(1), 158–171. https://doi.org/10.1037/1528-3542.7.1.158

Shephard, R. J. (2003). Limits to the measurement of habitual physical activity by questionnaires. *British Journal of Sports Medicine, 37*(3), 197–206. https://doi.org/10.1136/bjsm.37.3.197

Skiendziel, T., Rösch, A. G., & Schultheiss, O. C. (2019). Assessing the convergent validity between the automated emotion recognition software Noldus FaceReader 7 and Facial Action Coding System Scoring (J. De Houwer, Ed.). *PLOS ONE*, *14* (10), e0223905. Doi: 10.1371/journal.pone.0223905.

Teixeira, T., Wedel, M., & Pieters, R. (2012). Emotion-induced engagement in Internet video advertisements. *Journal of Marketing Research, 49*(2), 144–159. https://doi.org/10.1509/jmr.10.0207

Telpaz, A., Webb, R., & Levy, D. J. (2015). Using EEG to Predict Consumers' Future choices. *Journal of Marketing Research, 52*(4), 511–529. https://doi.org/10.1509/jmr.13.0564

Tian, Y.-L., Kanade, T., & Cohn, J. F. (2001). Recognizing Action units for Facial Expression Analysis. *IEEE transactions on pattern analysis and machine intelligence, 23* (2), 97–115. https://doi.org/10.1109/34.908962

Tkalčič, M., Maleki, N., Pesek, M., Elahi, M., Ricci, F., & Marolt, M. (2019). Prediction of music pairwise preferences from facial expressions. In *Proceedings of the 24th International Conference on Intelligent User Interfaces* (pp. 150–159). https://doi.org/10.1145/3301275.3302266

Tuan Pham, M. (2004). The Logic of feeling. *Journal of Consumer Psychology, 14*(4), 360–369. https://doi.org/10.1207/s15327663jcp1404 5

Venables, W. N., & Ripley, B. D. (2002). *Modern applied statistics with S* (4th ed.).Springer. https://www.stats.ox.ac.uk/pub/MASS4/.

Venkatraman, V., Dimoka, A., Pavlou, P. A., Vo, K., Hampton, W., Bollinger, B., Hershfield, H. E., Ishihara, M., & Winer, R. S. (2015). Predicting Advertising Success beyond Traditional measures: New Insights from Neurophysiological Methods and Market Response Modeling. *Journal of Marketing Research, 52*(4), 436–452. https://doi.org/10.1509/jmr.13.0593

Wang, Y. J., Minor, M. S., & Wei, J. (2011). Aesthetics and the online shopping environment: Understanding consumer responses. *Journal of Retailing, 87*(1), 46–58. https://doi.org/10.1016/j.jretai.2010.09.002

Wichmann, J. R., Uppal, A., Sharma, A., & Dekimpe, M. G. (2022). A global perspective on the marketing mix across time and space. *International Journal of Research in Marketing, 39*(2), 502–521. https://doi.org/10.1016/j.ijresmar.2021.09.001

Wilkinson, D., & Birmingham, P. (Eds.). (2011). *Using Research Instruments: A Guide for Researchers*. RoutledgeFalmer.

Yadava, M., Kumar, P., Saini, R., Roy, P. P., & Prosad Dogra, D. (2017). Analysis of EEG signals and its application to neuromarketing. *Multimedia Tools and Applications, 76* (18), 19087–19111. https://doi.org/10.1007/s11042-017-4580-6

Yoon, C., Gonzalez, R., & Bettman, J. R. (2009). Using fMRI to inform marketing research: Challenges and opportunities. *Journal of Marketing Research, 46*(1), 17–19. https://doi.org/10.1509/jmkr.46.1.17

Pantelis Karapanagiotis is Postdoctoral Researcher in Microeconometrics at the University of Groningen and an affiliate researcher at Leibniz Institute SAFE in Germany. He studied mathematics, and economics, and received his Ph.D. in quantitative economics from Goethe University Frankfurt. His primary research areas include structural industrial organization, computational economics, and behavioral decision theory.

Franziska Krause is Associate Professor for Marketing & Customer Insight at EBS University for Business and Law in Germany. She received her doctoral degree from Vienna University for Economics and Business in Austria. Her primary research interest lies at the intersection of consumer behavior and technology.

Janina Krick is a research assistant and doctoral student at the Chair of Customer Behavior and Sales at EBS University in Germany. She holds a BSc in General Management and a MSc in Management from EBS University. Her research is focused on hyper-personalization and consumeŕs perceptions of innovations in the digital realm.