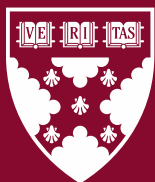


Working Paper 23-038

Summarizing the Mental Customer Journey

Julian De Freitas
Ahmet Uğuralp
Zeliha Uğuralp
Pechthida Kim
Tomer Ullman



**Harvard
Business
School**

Summarizing the Mental Customer Journey

Julian De Freitas
Harvard Business School

Ahmet Uğuralp
Bilkent University

Zeliha Uğuralp
Bilkent University

Pechthida Kim
Harvard College

Tomer Ullman
Harvard University

Working Paper 23-038

Copyright © 2023 by Julian De Freitas, Ahmet Uğuralp, Zeliha Uğuralp, Pechthida Kim, and Tomer Ullman.

Working papers are in draft form. This working paper is distributed for purposes of comment and discussion only. It may not be reproduced without permission of the copyright holder. Copies of working papers are available from the author.

Funding for this research was provided in part by Harvard Business School.

Summarizing The Mental Customer Journey

[Authors Anonymized]

CONSUMER RELEVANCE AND CONTRIBUTION STATEMENT

(WORDS: 286 [300 WORD LIMIT])

Life is filled with experiences that we undergo twice: once in the living and once in the recounting. When we go on a week-long vacation, we have the moment-by-moment experience of that vacation. But when a friend asks us how the trip was, we do not give them a week-long tale. Here we investigate how consumers summarize customer experiences, by creating variously patterned experiences and investigating which features of the patterns best predict these outcomes.

Our findings have several theoretical implications for work on summarization, architecting customer journeys, and predicting the success of content. Compared to previous work on summarization, we simultaneously consider a wider set of experience patterns and features of those patterns and quantify the relative importance of each feature in predicting consumer-relevant outcomes. Contributing to work on customer journeys, we help to empirically adjudicate between whether smooth or fluctuating customer journeys (Siebert et al. 2020) are more rewarding for customers. Finally, whereas previous work on predicting the success of content has considered features of the content itself (e.g., the words of a scientific paper), we quantify features of the mental experience (e.g., levels of enjoyment).

On a practical level, we show the merits of looking beyond a stage-wise conceptualization of customer journeys (Homburg, Jozić, and Kuehnl 2017) to focusing on their underlying patterns. Further, since managers have limited resources to invest in each touchpoint, we identify which aspects of the journey to focus investments on (the end, slope, and area under the curve). Finally, given potential customer backlash against collecting data on the journey

itself, we identify how firms can achieve similar results with possibly less invasive approaches, by leveraging sentiment scores of language customers use to describe their journeys.

ABSTRACT (WORDS: 198 [200 WORD LIMIT])

How do consumers summarize and act on their experiences, as when deciding whether an interaction with a firm was satisfying and whether to buy from it? Previous work on the summary of continuous experiences has tended to focus on a handful of experience patterns and features of those patterns, such as the area under the curve, peak value, and end value. Here, we consider a wider array of possible experience patterns and features of those patterns (27 patterns and 21 features) and quantitatively assess which features are most tied to consumer satisfaction and choice. Contrasting with theories that say fluctuating journeys are most effective, we find that consumers are most satisfied by journeys that stay consistently positive or improve over time, especially toward the experience's end. Furthermore, we find that *several* features of the experience predict these outcomes. These features include the end value, slope, integral, peak, and a sentiment score of the words people use to describe the experiences, although the consistently best predictor is the end value. The findings have theoretical implications for summarization, architecting customer journeys, and predicting the success of content, as well as practical implications for return on investment in customer experience optimization.

Keywords: customer experience; customer journey; customer satisfaction; natural language processing; summarization

Life is filled with experiences that we undergo twice: once in the living, and once in the recounting. When we go on a week-long vacation, we have the moment-by-moment experience of that vacation. But when a friend asks us how the trip was, we do not give them a week-long tale. Rather we summarize and compress the experience into a single evaluation, which affects what we recommend and what we ourselves consume. This tendency is prevalent in consumption, given that consumption often spans temporal intervals. When we visit the dentist or watch a horror film, we experience fluctuating levels of pain and unease. When we watch a performance such as a singer or interview a job candidate, we may be impressed at one moment and bored the next. When we interact with a company, we experience varying levels of satisfaction and annoyance over time. When we make the decision of whether to recommend a movie, to undergo a medical procedure again, or to buy a product, we somehow collapse and compress a continuous line of experiences into a summarized point.

This paper is about how consumers mentally compress their experiences. We focus especially on what *patterns* of mental experience lead to success, both when the experience is of a single offering (e.g., watching a movie trailer), and when it occurs across various touch points with a firm over a more extended time period, aka the ‘customer journey’ (e.g., as when you first see the firm’s ad, then log on to its website, receive information about the product, install the product, and so forth).

Previous works have examined how consumers summarize trajectories by studying a few experience patterns at a time, and by testing the importance of a few features at a time, such as peaks or velocity (Bhargave and Montgomery 2013; Carmon and Kahneman 1996; Fredrickson and Kahneman 1993; Hsee, Salovey, and Abelson 1994; Kahneman et al. 1993; Loewenstein and Prelec 1993). Yet there are very many possible consumer experiences, and we do not yet have a

comprehensive, empirically driven understanding of whether these patterns matter beyond a simpler stage-wise conceptualization of customer experiences like *pre-purchase*, *purchase*, and *post-purchase* (Bettman and Park 1980; Demmers, Weltevreden, and van Dolen 2020; Grewal and Roggeveen 2020; Hamilton et al. 2021; Schamp, Heitmann, and Katzenstein 2019) and, if so, which patterns are optimal.

Furthermore, even if patterns matter in customer experiences, managers must decide where to prioritize their limited resources. To this end, we also quantify several *features* of the patterns and of the language consumers use to describe them, in order to determine which features best predict consumer-relevant outcomes.

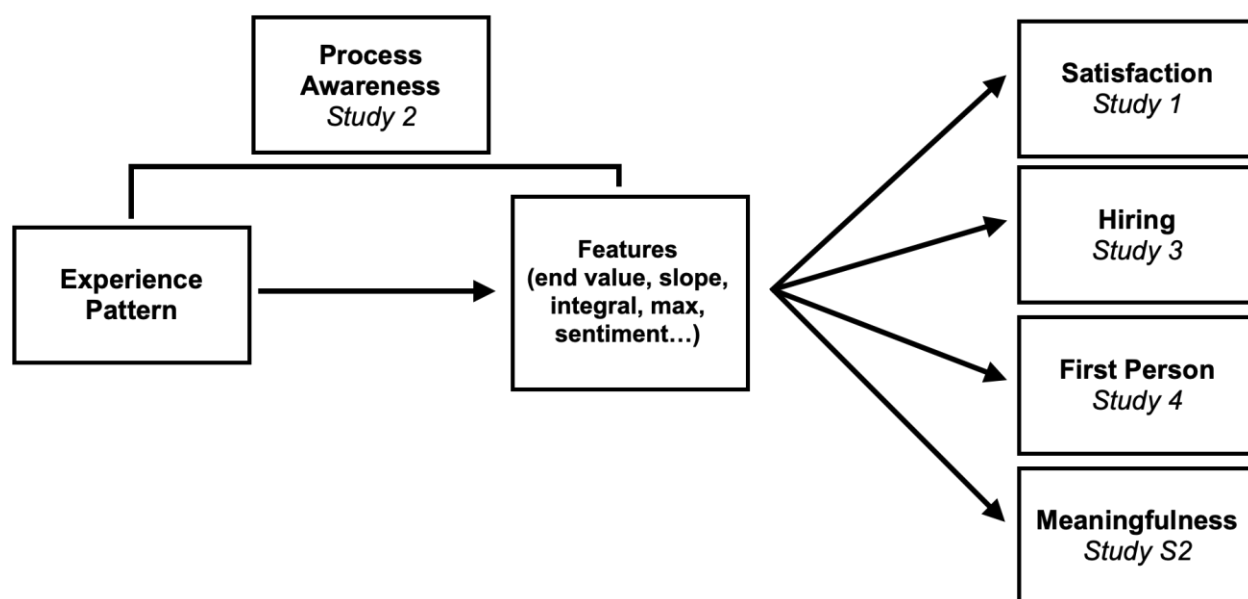
We then explore whether consumers are aware of how they summarize consumer experiences, or whether summarization is an automatic psychological processes that is informationally encapsulated from awareness (Bargh 2002; Bargh and Chartrand 1999b). If summarization is automatic, then marketers may influence it without consumers realizing how they did this.

Finally, we test whether summarization of consumer experiences generalizes across different experience domains at different temporal scales (e.g., a single offering, or an entire customer journey with a firm), as well as across first and third-person perspectives (Hung and Mukhopadhyay 2012), given that consumers make decisions not just based on how they summarize their own experiences but also based on how they summarizes the experiences of others.

In sum, the current work extends previous work in four major ways (figure 1): (i) it considers a wider range of experience patterns, by systematically constructing them and then quantifying which patterns are most successful; (ii) it assesses how well various features of these

patterns, both existing ones studied in previous work and new ones, predict consumer-relevant outcomes; (iii) it tests if consumers are introspectively aware of how they perform these summaries; and (iv) it tests if consumers summarize experience patterns similarly across both different *domains* of experience (e.g., a customer journey with a solar panel firm, or a single experience of a movie trailer), and across different *perspectives* (third versus first-person). Data, analysis code, and stimuli for all studies are available on [anonymized link].

FIGURE 1
CONCEPTUAL MODEL



Our findings have theoretical implications for work on summarization, architecting customer journeys, and predicting the success of content. Compared to previous work on summarization, we consider a wider set of experience patterns and features of those patterns, and assess the relative importance of each feature in predicting consumer-relevant outcomes. Contributing to work on customer journeys (Siebert et al. 2020), we help to empirically

adjudicate between whether smooth or fluctuating customer journeys are more successful.

Finally, whereas previous work on predicting the success of content has considered features of the content itself (e.g., the changing sentiment score of the words of a scientific paper), we study features of the mental experience (e.g., by measuring how consumers summarize their own continuous mental experiences, or summarize others' mental experiences).

On a practical level, we demonstrate the merits of moving beyond stage-wise conceptualizations of customer journeys (Demmers et al. 2020) to focusing on their underlying patterns. We also help identify which aspects of the journey managers should invest limited resources in (the end, slope, and area under the curve). Finally, given potential customer backlash against collecting data that reveals the customer journey, we identify how firms can achieve similar results with potentially less 'invasive' approaches, by leveraging sentiment scores of the language customers use to describe their journeys. In what follows, we outline our theoretical framework, followed by six experiments testing it.

THEORETICAL FRAMEWORK

What is the Optimal Pattern of a Customer Experience?

Managers deliver value not just through their offerings but also the overall customer experience (Schmitt 1999). Customer experiences occur at various scales, ranging from a single offering (e.g., a single movie trailer) to a longer 'customer journey' with a firm (i.e., an extended experience with a firm and its offerings that occurs across various touchpoints, such as from first viewing an advertisement all the way to re-purchasing the offering) (Edelman and Singer 2015;

Hildebrand and Schlager 2019; Lemon and Verhoef 2016). Despite the different temporal scales of these experiences, what they have in common is that they create *patterns* of mental experience in the customer over time, e.g., fluctuating levels of happiness, annoyance, or engagement. Firms have long measured consumers' continuous responses to their offerings, including their moment to moment reactions to television shows (Hui, Meyvis, and Assael 2014; Nelson, Meyvis, and Galak 2009), movies (Zhang, Wang, and Chen 2020), live streaming (Lin, Yao, and Chen 2021), and especially commercials (Aaker, Stayman, and Hagerty 1986; Baumgartner, Sujaan, and Padgett 1997; Hughes 1992; Madrigal and Bee 2005; Ramanathan and McGill 2007; Teixeira, Wedel, and Pieters 2012; Vanden Abeele and MacLachlan 1994; Woltman Elpers, Mukherjee, and Hoyer 2004; Woltman Elpers, Wedel, and Pieters 2003). The responses measured in these studies are quite varied, including emotional reactions, and ratings of liking, humor, and being entertained and informed.

Taking this dynamic view of the customer across time may help firms improve customer satisfaction, bolster sales, and grow customer lifetime value, as by acquiring new customers, extracting more margin from existing customers, and preventing churn (Schweidel et al. 2022). To these ends, most previous work on customer journeys has tended to organize them into discrete, stage-wise steps such as *pre-purchase*, *purchase*, and *post-purchase* (Homburg et al. 2017). These stages have then informed whether and what kind of data to collect, and what marketing activities to prioritize (Bettman and Park 1980; Demmers et al. 2020; Grewal and Roggeveen 2020; Hamilton et al. 2021; Schamp et al. 2019).

Moving from a stage-wise conceptualization to more of a continuous one, both practitioners and academics have begun to ask whether the *pattern* of the journey also matters and, if so, what kinds of patterns are most successful (Siebert et al. 2020). As of now, these

debates are mostly theoretical and qualitative, and has largely centered around whether managers may find more success from smooth journeys (Frow and Payne 2007; Kuehnl, Jozic, and Homburg 2019) or fluctuating journeys, which can be more effortful, inconsistent, and unpredictable (Alter 2017; Eyal 2014). Here, we empirically contribute to this debate by systematically generating a wide range of possible customer experience patterns (27 of them), and measuring whether these patterns affect consumer satisfaction, desire, and purchase intent. We predict:

H1 – The Pattern Matters. Consumer-relevant outcomes depend on the pattern of the experience.

What Features of These Patterns Matter?

If the continuous pattern of a customer journey matters, managers still need to decide how to prioritize investment of limited resources to gain visibility into these journeys and optimize them, e.g., devote human or other resources to improving a given touchpoint, like a landing page, payment processor, or customer service encounter.

Addressing this problem requires asking an additional question: *how* do customers summarize their experiences? If one thinks of an experience as tracing out a curve over time, one can quantify various features of that curve that could, in principle, inform how it is summarized. By quantifying all such features, one can test how well each feature performs at predicting consumer judgments and choices, which serves as an indicator of which features influence customers' summaries. Here, we explore a relatively exhaustive list of features inspired by, and

extending beyond, previous work (described below), and test their relative importance across a representatively large set of possible customer experience patterns.

In principle, many features can inform a customer's summary of an experience. One approach is 'sum up' the mental state(s) the customer experienced over the course of the experience, which is equivalent to 'plotting' their mental state (say happiness) over time and then mentally computing the area under the curve (Aaker et al. 1986; Thorson and Friestad 1989). While this approach is rational (i.e., the more area under the curve the better), many studies find that people are more inclined to weight some singular moments of the experience—such as the 'peak', 'trough', and 'end'—more heavily than others (Baumgartner et al. 1997; Diener, Wirtz, and Oishi 2001; Fredrickson and Kahneman 1993; Hui et al. 2014; Newman, Lockhart, and Keil 2010; O'Brien and Ellsworth 2012; Redelmeier and Kahneman 1996; Verhoef, Antonides, and De Hoog 2004). For instance, in one memorable study, participants undergoing a colonoscopy were asked to rate their pain in real-time. Surprisingly, the colonoscopy that induced pain for *longer* but which had a less intense peak and end was later evaluated as less aversive, suggesting that in retrospect participants remembered these singular moments best (Redelmeier and Kahneman 1996). When it comes to endings, consumers also proactively *choose* to end on a high note, e.g., they would rather listen to their less preferred song first, followed by their most preferred one (Kahneman, Ratner, and Kahneman 1997; Ratner, Kahn, and Kahneman 1999). Similarly, consumers prefer receiving bad news before good news (Legg and Sweeny 2014), and endings that are happy or familiar as opposed to sad or unfamiliar (Ross and Simonson 1991; Winet and O'Brien 2022).

Compared to endings, the impact of beginnings is more controversial. The literature on 'thin slicing' suggests that consequential outcomes are well predicted by first impressions, as

when predicting the popularity of political candidates (Todorov et al. 2005), musical performances (Tsay 2013), or predicting whether someone is morally culpable for a moral transgression (De Freitas and Hafri 2022). At the same time, studies that pit beginnings against endings, suggest that beginnings are less important than endings (Garnefeld and Steinhoff 2013; Harman and Porter 2021).

Beyond the integral, peak, beginning and end, consumers' summaries are also sensitive to how the experience changes over time, including its velocity (Hsee and Abelson 1991), acceleration (Hsee et al. 1994), and improvement or deterioration (Bhargava and Montgomery 2013; Loewenstein and Prelec 1993). Further, the direction of change and magnitude of the slope is important, particularly in the second half of the experience (Ariely 1998; Hansen and Danaher 1999; Hsee and Abelson 1991). Specifically, people prefer experiences that improve over time, especially towards the end, such as decreasing discomfort or an increasing wage profile (Loewenstein and Prelec 1993).

Here, we stress test these insights by simultaneously comparing the importance of these features across a wide range of possible experiences, instead of taking a more piecemeal approach that focuses on one or two features at a time. Our set of features both borrows from and extends beyond these previous works, including the following: how much the pattern is improving or deteriorating over time (the first derivative, aka slope, and various weightings thereof), whether such changes are fast or slow (the second derivative, aka acceleration, and various weightings thereof), the peak, valley, start, and end values of the line, the integral (aka area under the curve), and the number of peaks, valleys, and total extrema (i.e., peaks and valleys) (figure 2). We reason that, when it comes to summarizing a wide array of possible

experiences, more than one feature can be informative of whether the trajectory is going well or poorly overall. Thus, we predict:

H2 - Several Predictive Features. Several features of an experience predict consumer-relevant outcomes such as satisfaction and consumption choices.

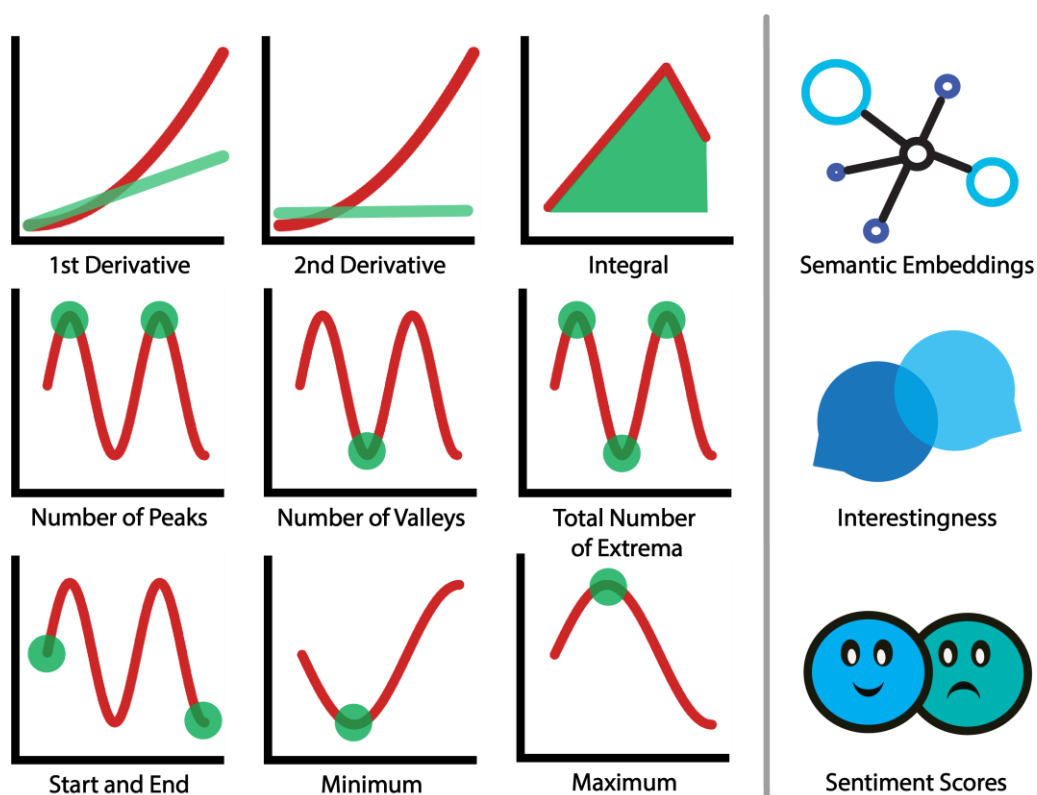
Yet, even if a manager knows which aspects of an experience to improve, he or she needs to collect the continuous data to inform these improvements, such as customers' search results, connections on social media, or geographic coordinates. Given potential customer backlash against collecting such data, we also identify how firms can achieve similar results with the potentially less 'invasive' approach of analyzing features of language. Previous work has predicted the success of language-based content—such as stories, academic articles, social media posts, songs, and movies—using features of the language content itself, such as speed of semantic progression, variation in sentiment, and similarities among content (Berger and Packard 2018; Laurino Dos Santos and Berger 2022; Reagan et al. 2016; Toubia, Berger, and Eliashberg 2021). In some cases, these features have been interpreted as a 'proxy' of mental experience, e.g., the speed of semantic progression could serve as an indicator of how exciting or engaging the content was for the reader (Berger, Kim, and Meyer 2021).

Here, we quantify similar linguistic features, but of the language which consumers use to describe their own consumers experiences and that of others. Based on these descriptions, we quantify several linguistic features (figure 2), including: how interesting the experience is (as represented by the number of unique words participants use to describe it), the semantic meaning of these words (as generated from a trained neural network), and their valence (as calculated by a

sentiment model). We treat interestingness as a proxy of engagement, and semantic embeddings as a reflection of the narratives accompanying certain experience patterns (Deighton 1992; Deighton, Romer, and McQueen 1989). Embeddings have also been fruitfully utilized to understand dynamic consumption patterns in previous work, such as the news content people consume over the span of years (Dhillon and Aral 2021). Our use of sentiment analysis is motivated by psychological work showing that consumers appear to represent various entities (e.g., other people, firms, or abstract concepts) in evaluative terms (De Freitas et al. 2018; De Freitas et al. 2022; Fazio et al. 1986; Knobe, Prasada, and Newman 2013; Osgood 1952).

FIGURE 2

MAJOR FEATURES USED AS PREDICTORS



NOTE.—We quantify both literal features of a trajectory (green) and semantic features of the language participants provide (blue).

Process Awareness and Generalizability

In line with work suggesting that some mental modules operate automatically without consumers being aware of the underlying psychological processes (Bargh and Chartrand 1999a; Kahneman 2011), we also ask whether consumers have introspective access into which features inform how they summarize customer experiences. Automatic psychological processes operate obligatorily, outside of deliberate intention and control (De Freitas and Alvarez 2018; Firestone and Scholl 2016; Logan and Cowan 1984). If summarization is automatic, then it is even more important that marketers be aware of how it operates, so that they can influence this process rather than attempting other ways of influencing how consumers summarize experiences. We operationalize introspective access by asking consumers to rank the extent to which different features influence their summaries. We predict:

H3 - Lack of process awareness: Consumers do not accurately rank the extent to which different experience features influence their summaries, suggesting that they are unaware of how they conduct these summaries.

Next, we test the extent to which hypotheses 1 and 2 generalize across several experience domains, including customer journeys related to buying a product (studies 1 and 4), consumption of interview performances (study 3), experiences of movie trailers (study 4), and even entire lives (study S2). Despite the myriad differences across these domains, we predict that consumers

largely summarize similar experience patterns across domains using the same psychological process:

H4 - Generalizability Across Domains: Across experience domains, consumers summarize different patterns of experiences similarly.

Finally, we test whether hypotheses 1 and 2 generalize across first and third-person experiences, i.e., regardless of whether one is summarizing another's experience (studies 1-3) or one's own (study 4). Both types of summaries are practically relevant, since consumers make consumption decisions based on how they summarize both their own experiences and that of others. We predict:

H5 – Generalizability Across Perspectives: Consumers summarize experiences similarly across first and third-person perspectives.

STUDY 1: SATISFACTION OF A CUSTOMER JOURNEY

How do people summarize an entire customer journey (Lemon and Verhoef 2016), given the happiness that the customer experiences over the course of that journey? Study 1 starts by measuring consumer intuitions about other's experiences, which can influence whether consumers consume these experiences themselves. Specifically, we explore what types of mental customer journeys are viewed as most satisfying and whether consumers personally desire such journeys for themselves. Note that, although it is possible for satisfaction and desirability to perfectly correlate, they may also come apart. For instance, even though a tumultuous journey

can be effortful, challenging, and ultimately satisfying, whether consumers desires such journeys for themselves may depend on their personal tolerance for adversity.

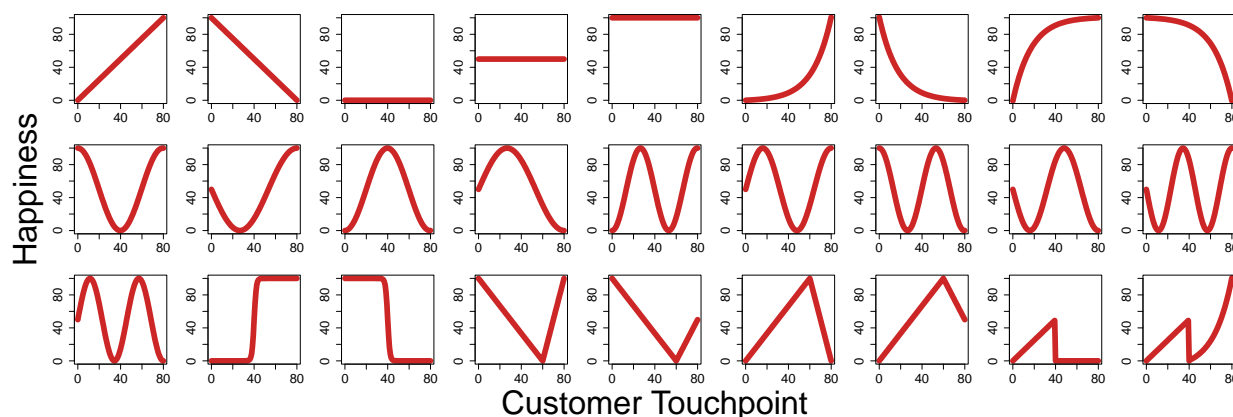
Method

Materials. We generated 27 mental customer journey lines, each of which depicted a fictional customer's experience with a solar panel firm named *Solaro*. The customer's "Happiness" was plotted on the y-axis across each "Customer Touchpoint" of the x-axis. The instructions explained what the touch points represented (more below).

To capture a wide range of customer journey patterns, we generated them from several basis functions (figure 3): linear (positive, negative, and constant slopes), exponential (various reflections), sinusoidal (various transformations), and logistic (jump and drop). We also included two custom-made trajectories inspired by the writer Kurt Vonnegut's "Old Testament" and "New Testament" story arcs (Vonnegut 1995): a linear rise followed by a sharp fall and a constant low, and a linear rise followed by a sharp fall and upward exponential rise (figure S1). The remaining Vonnegut plots were already captured by our existing functions.

FIGURE 3

CUSTOMER JOURNEY LINES IN STUDY 1



Procedure. We recruited 295 participants on Amazon’s Mechanical Turk who passed four attention checks and finished the survey in exchange for \$2.50. Since a pilot study showed that we would exclude around a third of participants due to stringent comprehension checks about survey content and interpreting basic graphs, we recruited 295 participants so that we would have at least 100 participants after exclusions. We ended up excluding 118 based on comprehension checks presented at both the beginning and end of the survey (described below), leaving 177 ($M_{age} = 40.5$, 46% female). Participants were given the following study instructions:

“In this experiment, we will show you the ‘customer experience lines’ of different customers who interacted with a solar panel firm called *Solaro*. On the y-axis, we will plot how the customer felt throughout their customer journey, and on the x-axis, we will plot each ‘touchpoint’ they had with *Solaro* during this journey. A ‘touchpoint’ is a moment when a customer interacts with the company.

Each customer had 80 touchpoints, from the point of first hearing about *Solaro* to eventually buying a solar panel from them. Examples of customer touchpoints include: reading their first *Solaro* ad, logging on to *Solaro*’s website, receiving information from *Solaro* about their current energy usage, obtaining estimates for how much it would cost to install solar panels on their roof, and so forth.

Therefore, the overall customer experience line shows how the customer felt at each touchpoint along their customer journey with *Solaro*.”

At this point, they were shown an example plot with ‘stress’ plotted on the y-axis across time on the x-axis, and asked three comprehension questions about it:

“At what customer touchpoint was the person above when they felt the most stressed in their experience? [0, 20, 40, 60, 80]”

“How stressed did the person above feel when they were at the 20th customer touchpoint? [0, 20, 40, 60, 80, 100]”

“Which is true of the customer experience of the person above?” [“They were highly stressed early in their customer experience, then highly unstressed later in their customer experience; They were highly unstressed early in their customer experience, then highly stressed later in their customer experience”; They were highly stressed both early in their customer experience, then highly stressed later in their customer experience”; They were highly unstressed both early in their customer experience and later in their customer”]

If participants failed any of these checks, they were shown a short educational video explaining how to interpret the axes of a graph. They were then shown a new graph and asked three similar comprehension questions as the ones above. Participants were excluded for failing any of these questions.

They were then shown a picture with thumbnails of all 27 customer journey lines that they would rate in the main study. Next, they were shown each of the 27 customer journeys on its own page, accompanied by the following four questions:

“How **satisfying** was this person’s customer experience **overall**?” (100-point slider, with 0 = Least satisfying customer experience possible and 100 = Most satisfying customer experience possible)

“How much would you like for **your** customer experience to look like this?” (100-point slider, with 0 = Not at all and 100 = Very much)

“If you had to summarize this person’s customer experience using just **one word**, what would it be?” (We presented a small text box)

Imagine that you wanted to install solar panels on the roof of your home, and your customer experience looked like the one above. How much would you be **willing to pay** Solaro to install solar panels on your home (customers typically pay around \$20,000 for a similar service from other solar panel companies)? Please enter the amount below without the dollar sign or any commas. (We presented a small text box)

All depending variables were consumer-relevant outcomes, except for the summarization question, which was intended for our natural language analyses. At the end of the survey, we

asked three more comprehension checks about what was plotted on the y-axis, x-axis, and what type of questions were asked:

“You just saw many plots. What was labeled on the y-axis [Customer Touchpoint, Happiness, Satisfaction, Age]”

“What was labeled on the x-axis? [Customer Touchpoint, Happiness, Satisfaction, Age]”

“The first question after each plot asked you to assess the following about the person’s customer experience line: [Customer Touchpoint, Happiness, Satisfaction, Age]”

Participants were also excluded for failing any of these questions. Altogether, the relatively high percentage of exclusions (40%) likely reflects a math illiteracy issue with Mturk participants, which further underscores why it was important to include so many comprehension checks to ensure data quality. Finally, participants completed basic demographics items (political orientation, gender, ethnicity, age, education level) and provided any comments on the survey.

Results

Descriptive Analyses. We treated customer journey type as a continuous variable ranging from 1 to 27, ordered according to average satisfaction scores. To test whether the pattern of a customer journey affected participant outcomes, we ran a mixed-effects linear regression, with question type (satisfaction or desirability) and pattern (of the various customer journey lines) as fixed factors and participant number as a random intercept. We found significant effects of question type ($b = -10.27, p < .001$) and customer journey type ($b = 2.58, p < .001$; figure 4), supporting hypothesis 1 that pattern matters. We also found that both satisfaction and personal desirability were significantly correlated with willingness to pay (satisfaction: $r = .42, p < .001$; personal desirability: $r = 0.39, p < .001$).

FIGURE 4

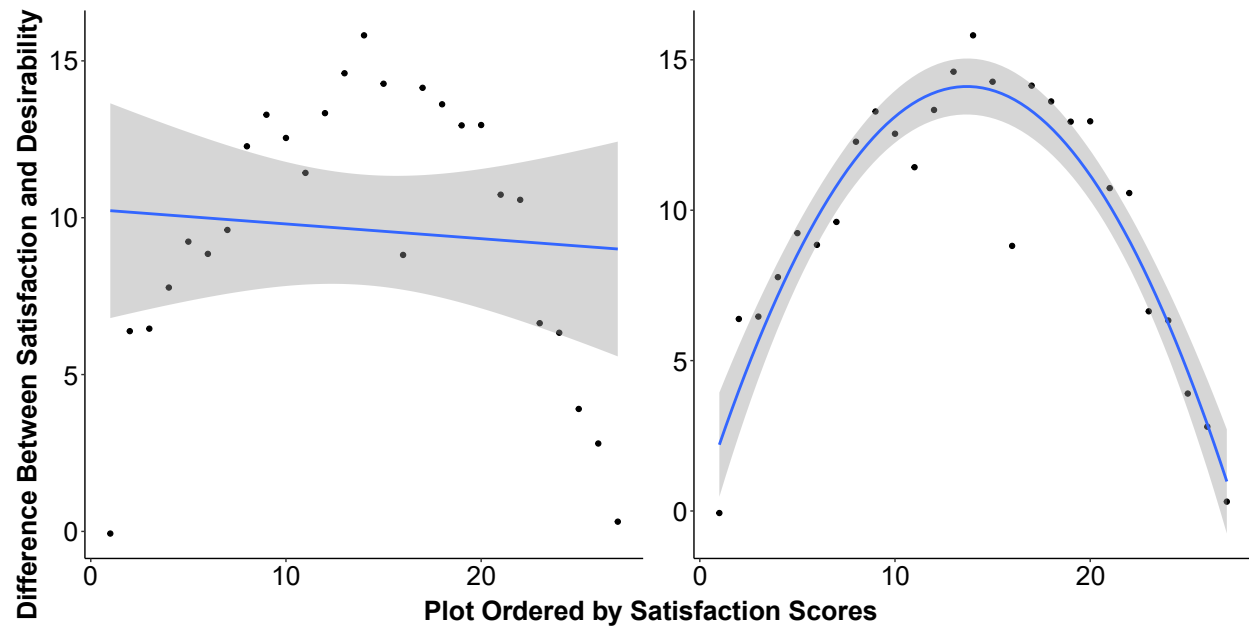
AVERAGE SATISFACTION AND DESIRABILITY OF CUSTOMER JOURNEYS, RANKED BY
ASCENDING SATISFACTION



Exploratory Analyses. We also noticed an interesting pattern between summaries of satisfaction versus desirability: When we plotted satisfaction and desirability scores arranged in ascending order of satisfaction, the differences between these outcomes were initially small, then increased, then decreased again (figure 4). Indeed, when we regressed the difference between satisfaction and personal desirability on customer journey numbers ordered by the average satisfaction scores, we found that a linear model fit the data poorly ($b = -0.05$, $p = .672$), whereas a quadratic model fit them well (quadratic: $b = -0.07$, $p < .001$; quadratic $>$ linear, $\chi^2(1) = 57.50$, $p < .001$; figure 5). Thus, satisfaction and personal desirability ratings only converged for customer journeys that were either extremely satisfying or unsatisfying, implying that people may avoid fluctuating experiences that they otherwise deem to be relatively more satisfying.

FIGURE 5

LINEAR AND QUADRATIC FITS TO THE DIFFERENCE BETWEEN SATISFACTION AND PERSONAL DESIRABILITY



NOTE.— Points are arranged by increasing satisfaction scores.

Predictive Analyses. So far, the results show that pattern matters (hypothesis 1). But how do consumers go from seeing a particular pattern to deciding how satisfying or desirable it is? To answer this question, we exhaustively quantified 18 *literal* features of the trajectories:

- *First derivatives*, with various weightings, i.e., ascending or descending across a customer journey, or weighting most heavily a line's beginning, end or "early" touchpoints. To calculate the ascending and descending weights, we divided the pattern into quarters, and assigned the quarters with weights of 0.25, 0.5, 0.75, and 1 in ascending or descending order. For beginning weights, we only considered the first quarter of the pattern, and for end weights, we only

considered the last quarter. For calculating the "early" weights, we only considered part of the pattern between 18 and 30 along the x-axis.

- *Second derivatives*, i.e., the line's accelerations
- *Integral*, i.e., the summation under the line
- *Number of peaks, valleys, or both*
- *Maximum and minimum*
- *End*, i.e., the last y-value
- *Start*, i.e., the first y-value.

We also included three *conceptual* features related to participants' word descriptors:

- *Sentiment score*, i.e., valence, representing evaluation, as computed using a deep learning-based sentiment model that was trained on billions of texts such as news articles and Wikipedia entries. The model uses the vector embeddings of words, which represent a word's meaning as a list of numbers, to calculate sentiment score, through the 'sentiment.ai' package in R.
- *Interestingness*, i.e., the count of unique words used to describe a given experience.
- *Semantic embedding*, a representation of each word's semantic similarity to other words, or its overall uniqueness, as learned by an artificial deep neural network that was trained to perform a variety of natural language tasks (Cer et al. 2018).

We then evaluated each of the 21 features separately. To quantify how well each feature predicted summaries of satisfaction and personal desirability, we used a k-fold cross-validation procedure with the *k* value (i.e., the number of folds) set to ten, which provides a good balance between bias and variance (Hastie et al. 2009; Kohavi 1995). For each of the ten folds, we

trained a regression model on all the data except the held-out fold, for which we generated fold-size predicted ratings. We then correlated these predicted ratings with the true ratings participants provided in that fold. We repeated this procedure for each fold, so that for each feature we had ten correlation values, one for each fold. We repeated this whole process for ten runs, to make sure our model was robust. A feature's final performance was simply the arithmetic mean of the hundred correlations (10 folds x 10 runs).

In line with hypotheses 2 and 3, summaries of satisfaction and personal desirability were significantly predicted by several features, both literal and linguistic, with the end value and sentiment score coming out on top (table 1; figure 6).

TABLE 1
MEAN PERFORMANCE OF STATISTICALLY SIGNIFICANT PREDICTORS IN STUDY 1

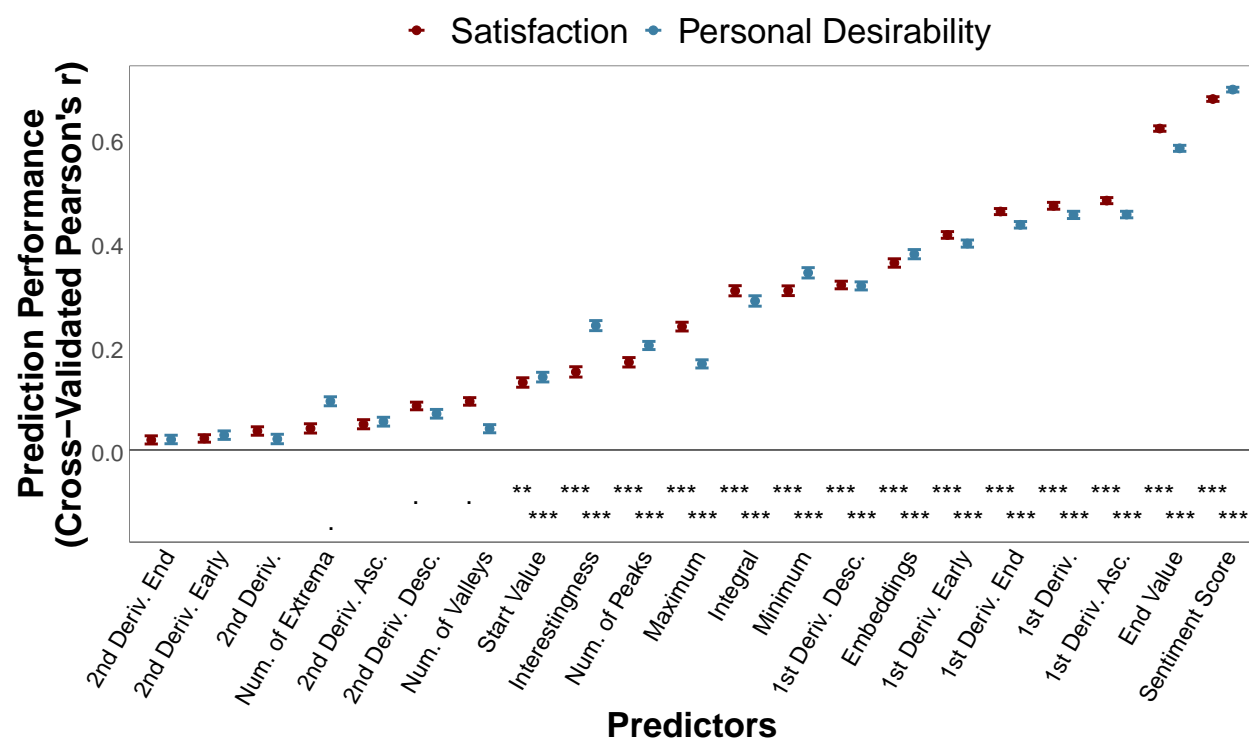
Predictor	<i>r</i> Satisfaction	<i>r</i> Desirability
Sentiment score	0.68 ^{***}	0.70 ^{***}
End value	0.63 ^{***}	0.59 ^{***}
First derivative (asc. weights)	0.49 ^{***}	0.46 ^{***}
First derivative (unweighted)	0.48 ^{***}	0.46 ^{***}
First derivative (end weights)	0.47 ^{***}	0.44 ^{***}
First derivative (early weights)	0.42 ^{***}	0.40 ^{***}
Embeddings	0.37 ^{***}	0.38 ^{***}
First derivative (desc. weights)	0.32 ^{***}	0.35 ^{***}
Minimum	0.31 ^{***}	0.32 ^{***}

Integral	0.31 ^{***}	0.29 ^{***}
Maximum	0.24 ^{***}	0.17 ^{***}
Number of peaks	0.18 ^{***}	0.21 ^{***}
Interestingness	0.16 ^{***}	0.25 ^{***}
Start value	0.14 ^{**}	0.15 ^{***}

NOTE.— Significant predictors ranked best to worst according to satisfaction. Predictors not featured were not statistically significant. ** = $p < .01$, *** = $p < .001$.

FIGURE 6

PERFORMANCE OF FEATURES IN PREDICTING SATISFACTION AND PERSONAL DESIRABILITY IN
STUDY 1



NOTE.— Circles depict the mean prediction performance, and error bars depict 95% CI. Features are ranked by the mean of the cross-validated Pearson's r value of satisfaction (red). Horizontal line and gray shadings depict mean and standard deviation of the prediction performance of a feature with randomly set values for each pattern. Stars beneath boxes indicate whether the root mean squared prediction errors are significantly less than that of a feature with random values. If the sample distribution was

normal, we used a Welch two sample t-test, otherwise, we used a Wilcoxon signed-rank test. '!' = $p < .1$, *** = $p < .001$. Lack of stars indicate non-significance.

We also corroborated the novel finding about sentiment score in three ways. First, we created word clouds of the word descriptors and noticed that the highest frequency words for each plot were indeed evaluative, e.g., the most frequent words were *perfect*, *great*, *good*, *satisfied* for the top five most satisfying customer journeys, and *horrible*, *bad* for the five least satisfying ones (figure S2). Second, using data-driven clustering of the patterns based on natural associations among the words used to describe them (aka topic models; Blei, Ng, and Jordan 2003), we found clusters intuitively corresponding to fluctuation, positive valence, and negative valence (figure S3). Third, and finally, we ran a follow up study to verify that the evaluative language participants employed was not simply ‘contaminated’ by the fact that they were being asked to make ratings of an evaluative nature (i.e., about satisfaction and desirability) on the same page. In study S1, participants were asked to *only* describe each experience with a word, without also rating it on satisfaction and desirability. We found that sentiment scores of these words correlated strongly with those of study 1, suggesting that the evaluative descriptions provided in study 1 were spontaneous rather than contaminated by the rating items.

Discussion

How do consumers judge whether a customer journey is satisfying, given its pattern? In line with hypothesis 1, we noticed that ratings of customer satisfaction and desirability were most positive for patterns that improved, followed by fluctuations, then followed by deteriorations. Interestingly, participants also thought that fluctuating journeys were more

satisfying than they were desirable. Perhaps, they recognized that tumultuous experiences can be satisfying, yet preferred to avoid adversity for themselves.

In line with the intuitive ordering of patterns and hypothesis 2, we also found that judgments were predicted by several literal features of the pattern rather than just one, including (from least to most predictive) its start value, interestingness, number of peaks, maximum, integral, minimum, semantic embeddings, various weightings of the first derivative (especially with ascending weights, indicating that later moments are more important), end value, and (the best predictor of all) sentiment score.

These results reveal the novel role of language sentiment. They also corroborate the importance of endings that has been found in previous work (Diener et al. 2001; Fredrickson and Kahneman 1993; Newman et al. 2010; O'Brien and Ellsworth 2012; Redelmeier and Kahneman 1996), but while doing so across a wider set of customer experiences and when predicting consumer-relevant outcomes like customer satisfaction and willingness to pay. Equally interesting is that acceleration was not a significant predictor. This result suggests that, while consumers care about whether an experience improves or deteriorates over time, they do not consider whether the acceleration of this improvement/deterioration is positive, negative, or zero (as can be seen for the most successful patterns depicted in figure 4).

Finally, the high performance of the sentiment predictor suggests that the language consumers use to summarize experiences provides a viable alternative approach for firms to gain insight into customer journeys when they or their consumers prefer not to collect continuous data on the customer journey itself.

STUDY 2: PROCESS AWARENESS

Study 1 found which are the most and least significant predictors of satisfaction and personal desirability. To what extent are participants introspectively aware of the relative importance of these features in informing their summaries of others' experiences?

Method

Materials. We used the same 27 customer journeys from study 1.

Procedure. We recruited 100 participants on Amazon's Mechanical Turk who passed the attention checks and finished the survey, in exchange for \$1.88. Since this time our analysis was simply to find the optimal ranking of features participants *thought* would predict satisfaction, we anticipated that a third of the subjects from study 2 would be sufficient. We excluded 34 based on the same stringent comprehension checks as study 1 at the beginning of the study and three new checks at the end of the study (described below), leaving 66 ($M_{\text{age}} = 38.3$, 59% female).

To ensure that participants understood the experimental paradigm from study 1, they first read the same instructions of that study, including seeing a picture of all the customer journey lines from that study. Then, to ensure that participants had a concrete understanding of that study, they answered the same four questions from study 1 for just one of the customer journey lines (the plot in which there is a linear rise, followed by a sharp fall and exponential rise). Next, they read the following:

“Imagine that you rated all 27 lines, and so did 200 other people who participated in this experiment. Which of the following factors do you think would be the best indicators of the ratings people provided?” Please arrange them from most to least useful, by clicking and dragging them.”

We presented icons of the 12 main predictors (excluding the various weightings of the first and second derivative predictors) accompanied by labels and intuitive descriptions of each, e.g., “Slope: The steepness of a line”. The predictors were initially arranged in random order, with the topmost predictor numbered #1 (most useful) and bottommost numbered #12 (least useful). We asked two comprehension questions about the ranking procedure and list of features:

“You just ranked many features. Which rank was for the most important feature? [1, 12, I don’t remember.]”

“Which of the following was *not* on the list of features? [Slope, Number of Valleys, Number of X-Values, Sentiment Scores]”

We excluded participants who failing any of these comprehension checks. Finally, we listed all features again and asked participants to select any that they did not understand. We excluded participants who did not understand more than two of the features. The feature that fewest participants understood was semantic embeddings (although 75% still understood it), while all remaining features were understood by more than 90% of participants. Participants then completed the same demographics items from study 1.

Results

To determine the optimal ranking of predictors given participants’ responses, we conducted rank order analysis using the Monte Carlo cross-entropy algorithm, which uses a distance criterion to find the optimal rank given a combination of different rankings (Pihur, Datta, and Datta 2007). To check that this method conformed with our assumptions, we also sorted the features by their mean rankings, and found a very similar aggregated ranking to the

Monte Carlo cross-entropy algorithm solution, with the only difference being that the minimum and number of valleys switched rankings. Importantly, we found that the optimal ranking from this analysis was different from the true rank found in study 1 (table 2). When correlating the predicted ranks from study 2 with the true ranks of study 1, we saw no evidence of a systematic relationship, $r = -0.11$, $p = .729$ (figure 7). If anything, consumers systematically underestimated features that were important, and overestimated features that were unimportant (table 2).

FIGURE 7

PREDICTED VERSUS TRUE RANKINGS OF PREDICTORS IN STUDIES 1 AND 2

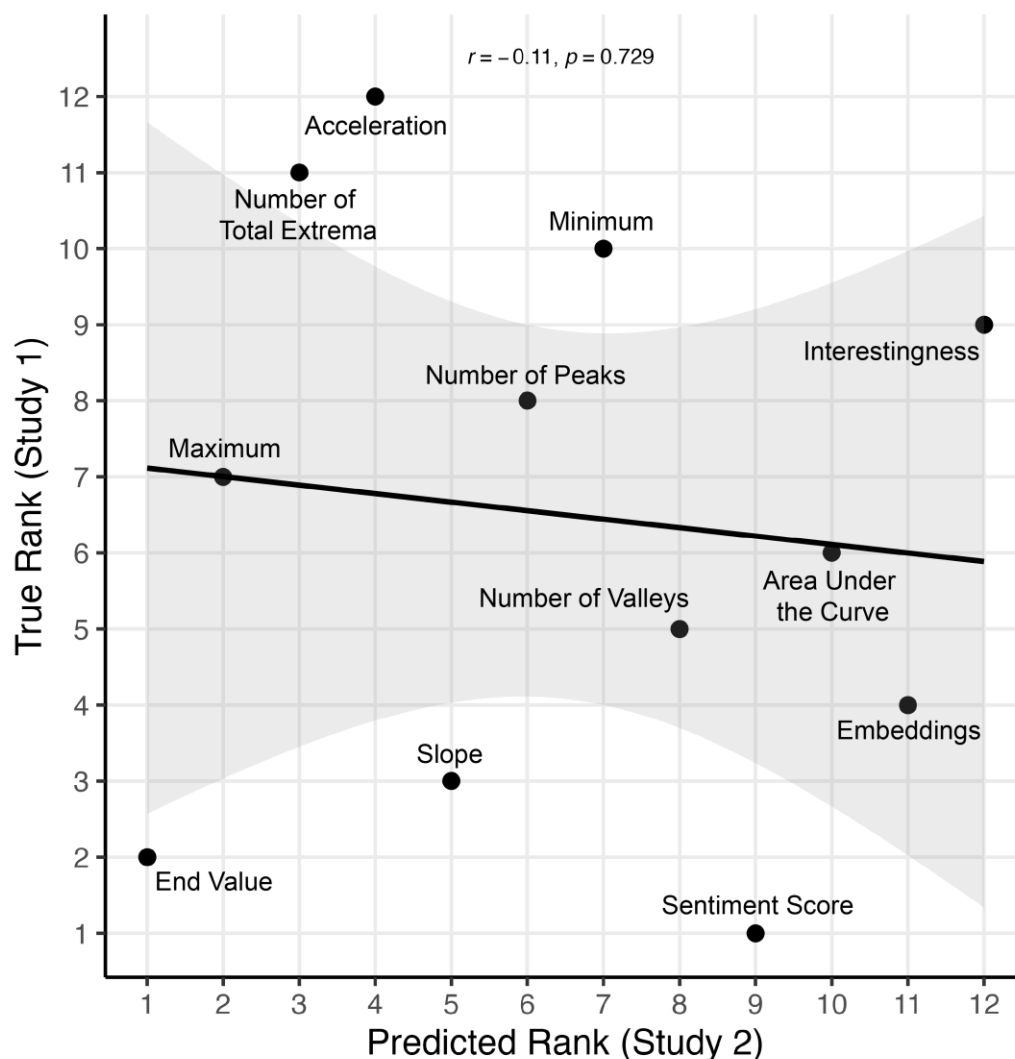


TABLE 2

PREDICTED VERSUS TRUE RANKINGS OF PREDICTORS IN STUDIES 1 AND 2

Predicted Rank (Study 2)	True Rank (Study 1)
End value (+1)	Sentiment score
Maximum (+5)	End value
Number of total extrema (+8)	Slope
Acceleration (+8)	Semantic embeddings
Slope (-2)	Minimum
Number of peaks (+2)	Area under the curve
Number of valleys (+3)	Maximum
Minimum (-3)	Number of peaks
Sentiment score (-8)	Interestingness
Area under the curve (-4)	Number of valleys
Semantic embeddings (-7)	Number of total extrema
Interestingness (-3)	Acceleration

NOTE.— Parenthetical numbers in left column indicated whether participants over or under-estimated the importance of a given feature relative to its true rank in study 1 (right column).

Discussion

Supporting hypothesis 3, the ranking of predictors generated by participants did not match the true order of predictors found in study 1, suggesting that participants were

introspectively unaware of the psychological processes underlying their summaries. If anything, they tended to place more importance on predictors that were less important in practice, e.g., sentiment score was expected to be one of the least important features when in practice it was the most important, and acceleration was expected to be one of the most important when in practice it was the least important. In short, the order of predictors found in study 1 is unintuitive and surprising.

STUDY 3: GENERALIZATION ACROSS EXPERIENCE DOMAINS: PERFORMANCE CONSUMPTION

How general are the findings of study 1? In this study we begin to consider the possibility that the previous findings generalize to entirely different experience domains, by investigating the case study consuming performances (Deighton 1992). Specifically, we focus on interview performance and the outcome of hiring likelihood. Interview performances differ from customer journeys in that an interviewee promotes themselves, typically within one session rather than over several touchpoints, and it is the interviewer(s) who must summarize their impression of the interviewee, as by deciding whether to hire the candidate after experiencing their job talk.

Most previous work on hiring decisions has focused on the role of more ‘static’ factors, such as a candidate salesperson’s appearance, personality, gender, and socioeconomic status (Olian, Schwab, and Haberfeld 1988; Sharps and Anderson 2021; Shtudiner 2019), and features of the interviewer such as overconfidence (Kausel, Culbertson, and Madrid 2016). Yet, as with customer journeys, one’s experience with a candidate traces out a pattern of mental states over

time, raising the possibility that a similar summarizing process to that uncovered in study 1 is at play and affects the consequential decision of whether someone is judged as likely to be hired.

Method

Materials. We created 27 interview performance lines, identical in pattern to the 27 customer journey lines of study 1. This time, the plots depicted a fictional candidate's perceived interview performance for a teaching position at a fictional university called *Northride College*, with "Perceived Performance" plotted on the y-axis and "Time" plotted on the x-axis.

Participants

Procedure. We recruited 296 participants on Amazon's Mechanical Turk who passed the attention checks and finished the survey, in exchange for \$2.50. We excluded 155 based on comprehension checks, leaving 141 ($M_{\text{age}} = 38.5$, 46% female). Attention checks and the various comprehension checks were similar to study 1, except that the answer options were adjusted to be about the hiring scenario. Participants were told the following:

"In this experiment, we will show you the 'interview performance lines' for different candidates who interviewed for a teaching position at a university called *Northride College*. On the y-axis, we will plot the perceived performance of the candidate throughout their interview, and on the x-axis, we will plot the time in minutes.

Each candidate was interviewed for 80 minutes. Therefore, the overall interview performance line shows the perceived performance of the candidate at each minute of their interview at *Northride College*."

Participants saw all 27 interview performance lines in a randomized order. Each interview performance line was presented on its own page with the following two questions:

“How likely is it that you would hire the candidate?” (100-point slider, with 0 = Extremely unlikely and 100 = Extremely likely)

“If you had to summarize this candidate’s interview performance using just **one word**, what would it be?” (We presented a small text box)

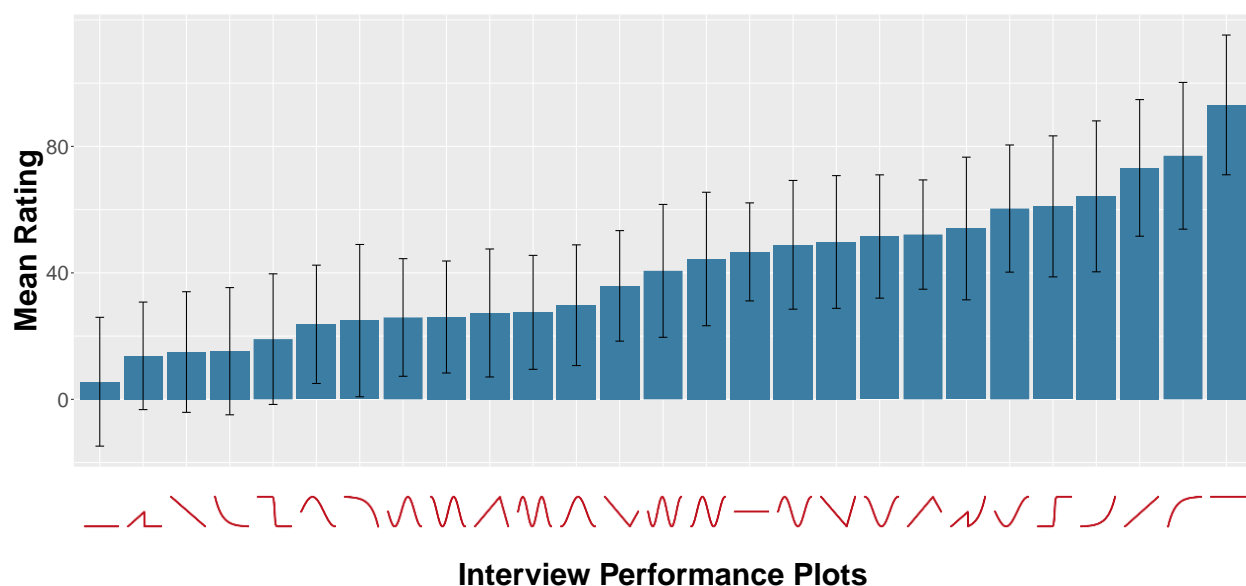
Participants then completed the comprehension checks and the same demographics items from study 1.

Results

Descriptive Analyses. We treated interview performance pattern as a continuous variable ranging from 1 to 27, ordered according to average hiring likelihood scores. To determine whether the pattern of the interview performance lines affected participant outcomes, we ran a mixed-effects linear regression with pattern as a fixed factor and participant number as a random intercept. We found a significant effect of pattern on hiring decisions, in line with hypothesis 1 ($b = 2.68, p < .001$; figure 8).

FIGURE 8

AVERAGE LIKELIHOOD OF HIRING, RANKED BY ASCENDING SCORES



Predictive Analyses. Using the same 21 predictors quantified in study 1, we utilized the ten-fold cross-validation approach to quantify how well each predictor predicted participant ratings (prediction accuracy = cross-validated Pearson's r). In line with hypothesis 2, summaries of hiring likelihood were significantly predicted by several features, with end value and sentiment score again coming out on top (table 3; figure 9). The order of the top five most successful predictors were almost identical to study 1, except that this time the unweighted first derivative performed slightly better than the first derivative with ascending weights.

TABLE 3

MEAN PERFORMANCE OF STATISTICALLY SIGNIFICANT PREDICTORS IN STUDY 3

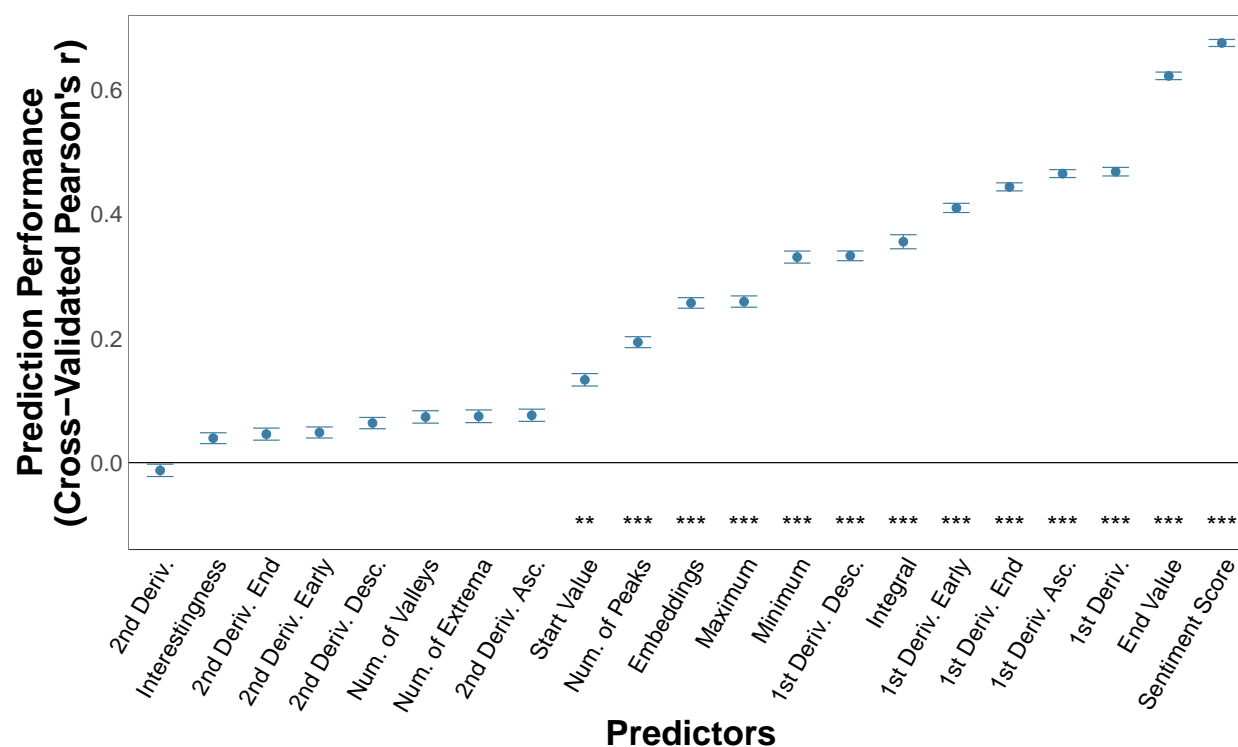
Predictor	r Hiring Likelihood
Sentiment score	0.67***
End value	0.62***

First derivative (unweighted)	0.47***
First derivative (asc. weights)	0.46***
First derivative (end weights)	0.44***
First derivative (early weights)	0.41***
Integral	0.36***
First derivative (desc. weights)	0.33***
Minimum	0.33***
Maximum	0.26***
Embeddings	0.26***
Number of peaks	0.19***
Start Value	0.13**

NOTE.— Significant predictors ranked best to worst. Predictors not featured were not statistically significant. ** = $p < .01$, *** = $p < .001$.

FIGURE 9

PERFORMANCE OF FEATURES IN PREDICTING HIRING LIKELIHOOD



NOTE.— Circles depict the mean prediction performance, and error bars depict 95% CI. Horizontal line and gray shadings depict mean and standard deviation of the prediction performance of a feature with randomly set values for each pattern. Stars beneath boxes indicate whether the root mean squared prediction errors are significantly less than that of a feature with random values. If the sample distribution was normal, we used a Welch two sample t-test, otherwise, we used a Wilcoxon signed-rank test. ** = $p < .01$, *** = $p < .001$. Lack of stars indicate non-significance.

As in study 1, we also corroborated the importance of sentiment scores by plotting the top five most and least satisfying word clouds, and again saw that the most frequent words for each trajectory were evaluative (e.g., *surprising*, *improving*, *impressive*, *perfect* for top five most satisfying interview performance lines and *terrible*, *poor*, *declining*, *disappointing* for the least satisfying ones; figure S4). We also corroborated the sentiment results with topic modeling, finding three clusters corresponding to negative, fluctuating, and positive interview performance trajectories (figure S5).

Discussion

In line with hypothesis 4, we found evidence that the same phenomenon uncovered when summarizing customer journeys (study 1) is at play in another practical domain involving personal performance marketing: hiring decisions. In line with hypotheses 1 and 2, hiring likelihood depended on the experience's *pattern*, with similar types of patterns (table 4) and predictors succeeding best.

Study S2 in the web appendix further explores generalization of the current summarization process, by testing whether the findings generalize even to a domain with a much longer temporal scale: summarizing the meaningfulness of entire lives. We find that similar literal features of a trajectory predict how meaningful a life is viewed as being, although the ranking of these features differ slightly from that for customer journeys and hiring likelihood. As in studies 1 and 2, we find that sentiment score is the best predictor, followed by the end value. But, unlike study 1 and 3, we also find that the integral and max are more predictive than the first derivative features, perhaps because the first derivative (aka slope) is viewed as less controllable by the agent across the longer span of their entire life.

Finally, we compared ratings of satisfaction (study 1), hiring (study 3) and meaningfulness (study s2) across studies, as well as ratings of desirability in the studies that measured it (studies 1 and s1), and found high correlations throughout (table 4). In line with hypothesis 4, these high correlations suggest that the same psychological process underlies consumers' summaries across these domains.

TABLE 4

CORRELATIONS BETWEEN RATINGS ACROSS STUDIES 1, 3, AND S1

Customer journey satisfaction v. personal desirability	0.89***
--	---------

Customer journey satisfaction v. life meaningfulness	0.87***
Customer journey satisfaction v. hiring	0.99***
Customer journey personal desirability v. life desirability	0.93***
Customer journey personal desirability v. hiring	0.98***
Life meaningfulness v. hiring	0.88***
Life personal desirability v. hiring	0.95***

STUDY 4: DIRECTLY EXPERIENCED CONTENT

Studies 1, 3, S1 and S2 considered how consumers summarize customer journeys experienced by others, as depicted by a trajectory over time. Study 4 explored whether these effects generalize to *first-person* experiences (hypothesis 5). To capture these experiences, we asked participants to indicate their continuous levels of enjoyment of movie trailers, using a procedure inspired by previous studies using continuous report (Aaker et al. 1986; Pham et al. 2001). We also leveraged an incentive-compatible design to test which experience patterns were most likely to lead to a consumption choice.

Method

Materials. We picked movie trailers that had not yet been released at the time of the experiment, drawn from 8 different genres: (1) Adventure: *Dungeons & Dragons: Honor Among Thieves*; (2) Action: *Mission: Impossible – Dead Reckoning*; (3) Animation: *Puss in Boots: The*

Last Wish; (4) SciFi: *Avatar: The Way of Water*; (5) Fantasy: *Shazam! Fury of the Gods*; (6) Biography: *I Wanna Dance With Somebody*; (7) Drama: *She Said*; and (8) Horror: *Knock at the Cabin*. To control the amount of content viewing time across trailers, we showed participants only the first 90 seconds of each trailer.

Procedure. We recruited 297 participants on Prolific who passed the attention checks and finished the survey, in exchange for \$3.30. Exclusions per attention checks were as in study 1. We excluded 81 participants based on comprehension checks and video checks (described below), leaving 216 ($M_{\text{age}} = 40.7$, 44% female). Participants were shown the following instructions, together with a video and an adjustable slider scale anchored from 0 (*Not enjoying*) to 100 (*Enjoying*):

“In this survey, you will rate your enjoyment throughout by continuously adjusting a slider under the video. For example, if you’re really enjoying one moment you can move the slider to the right, but if the next moment you are not enjoying it you can move the slider to the left.”

To ensure that participants understood the task, they were first presented with a demo video, which was a screen video recording of how one of the authors continuously adjusted the slider while watching a trailer for the movie, *Garfield* (2004). To ensure that participants understood the instructions, they were then asked to complete a practice trial in which they, too, watched the first 60 seconds of the *Garfield* trailer while indicating their continuous enjoyment levels. After completing this practice trial, participants were asked two comprehension questions about the trailer they saw and how they were asked to rate it: “Which movie was shown in the demo video?” (options: *Tom & Jerry*, *Rocky 2*, *Terminator 3*, *Garfield*), and “In which way were you asked to rate your enjoyment?” (options: *Continuously adjusting the slider throughout the video*, *Adjusting the slider only at the end of the video*, *Select it from a multiple choice question*).

Participants then completed the main trials, in which they saw each of the 8 movie trailers presented on its own page, with the trailers shown in randomized order. In addition to indicating their continuous levels of enjoyment during the video, after watching a video they indicated how willing they were to pay for the full movie, and to summarize the trailer using one word:

“How willing are you to pay to watch the full movie?” (100-point slider, with 0 = Extremely unlikely and 100 = Extremely likely)

“Please provide one word to describe the trailer.” (We presented a small text box)

We took several steps to ensure high data quality for the continuous enjoyment ratings. First, we collected participants’ enjoyment values every hundred milliseconds, so that we would have 900 points per trailer for each participant at the end of the survey. Second, we programmed the video interface so that participants were not able to fast forward the trailer or skip to the next screen until the trailer finished playing. Third, we reminded participants to keep reporting their enjoyment continuously. If the slider was stationary for more than 20 seconds, a warning message about this appeared above the trailer for 5 seconds while the video kept playing, before disappearing again. We excluded 58 participants who were inactive, i.e., who did not move the slider for more than 30 seconds. Fourth, we measured whether the quality of the video affected willingness to pay ratings, by collecting both the video resolutions in which the videos played (e.g., 144p, 240p, etc.), and the number of buffering interruptions that were experienced (i.e., each transition from a loading to a playing state). For each of the eight genres, we found that neither the resolution quality (all $ps > .05$) nor number of interruptions (all $ps > .05$) affected willingness to pay.

At the end of the survey, participants were informed that there was an opportunity to win a free ticket to watch the full movie of one of the trailers they saw:

“If you could watch only one of the following movies, which one would you watch?

Based on your choice, we will enter you into a raffle competition. If you win, you will receive a gift card to rent the movie.”

To remind the participants of the movies they watched, we presented cover art for each movie alongside the movie’s title, with the order of movies randomized between subjects, and participants could select just one movie. Participants then answered two comprehension checks about which movies they watched and what type of questions they were asked: “You just watched some trailers. Please select the one that you watched.” (options: *The Godfather*, *Jurassic World Dominion*, *Shazam! Fury of the Gods*, *Tom & Jerry*, *Pulp Fiction*)” and “Which of the following were you asked to rate? (options: *satisfaction*, *anticipation*, *enjoyment*, *boredom*). 23 participants were excluded for failing any of these questions and 58 were excluded based on the video checks, leading to 81 total exclusions. All participants completed the same demographics items from study 1. At the end of the study, we entered participants into the raffle and then awarded the winning participant.

Results

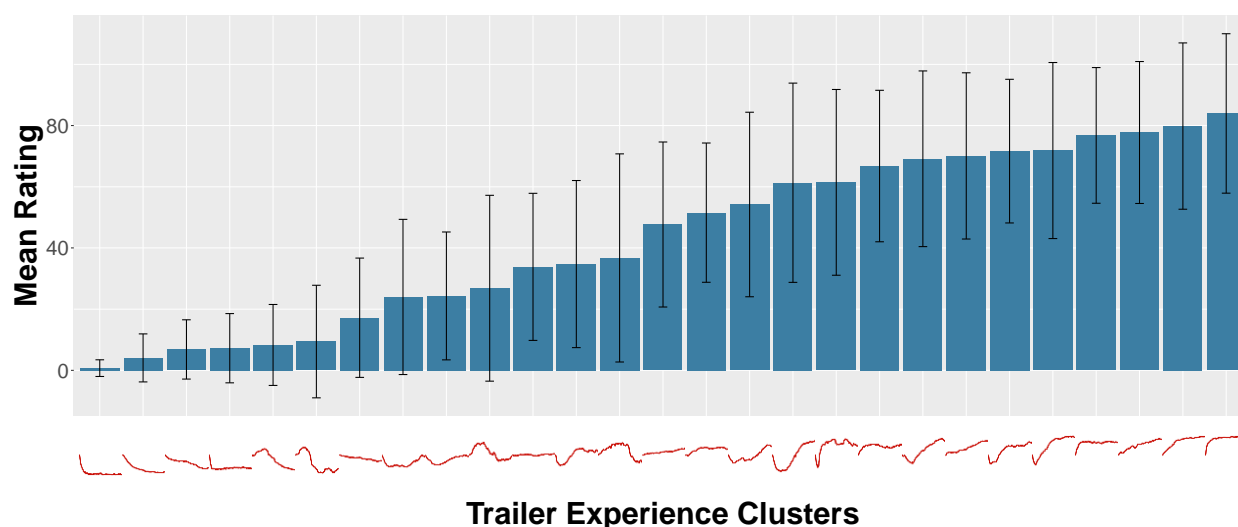
To convert the continuous experience data from each participant into an experience line, we fit their data using a least squares polynomial fit. After calculating the fitting errors of polynomial degrees ranging from 1 to 150 (figure S6), we picked the degree with the minimum average error, i.e., 68. 68 degrees is also around where the fit explains most of the error, i.e., it is near the elbow of the plot in figure S6. We used these fitted polynomials for subsequent descriptive and predictive analyses. Next, we used the k-means clustering algorithm, a method

commonly used to distinguish k number of different clusters from the data, to cluster the lines into common experience patterns (MacQueen 1967). To match the number of patterns presented in the previous studies, we generated 27 clusters in total (figure S7).

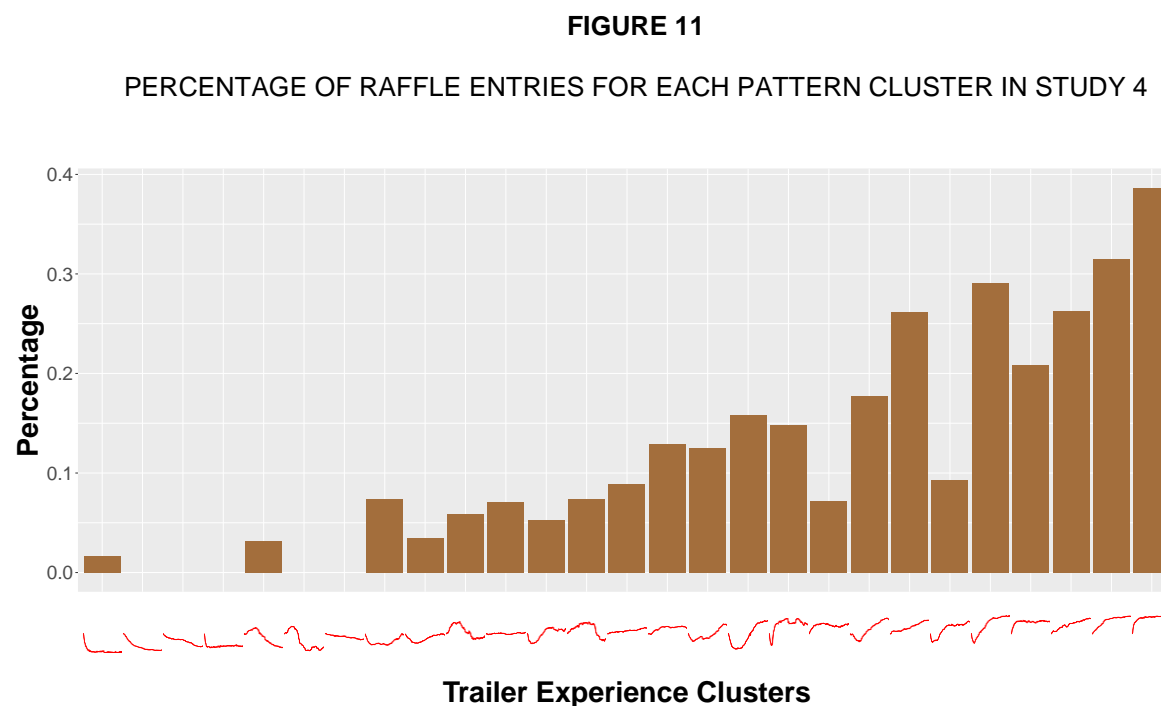
Descriptive Analyses. To determine whether the enjoyment patterns of participants affected willingness to pay, we ran a mixed-effects linear regression with cluster label as a fixed factor and participant number as a random intercept. We treated cluster label as a continuous variable ranging from 1 to 27, ordered according to average willingness to pay scores. We found a significant effect of cluster type on willingness to pay ($b = 3.37, p < .001$; figure 10). Strikingly, we also found that the ordering of the experience patterns intuitively resembled that of studies 1 and 3, with the leftmost cluster lines in figure 10 generally deteriorating, middle clusters fluctuating, and rightmost cluster lines improving. This ordering also related to which movies were chosen for raffle entry, since the percentage of raffle choices for each cluster correlated strongly with willingness to pay for the same clusters ($r = .87, p < .001$; figure 11).

FIGURE 10

AVERAGE WILLINGNESS TO PAY, RANKED BY ASCENDING SCORES, IN STUDY 4



NOTE.— Red lines under each bar indicates the average line of each cluster.



Predictive Analyses. We used the same 21 predictors quantified in study 1. Unlike in the descriptive analysis, we used each pattern generated by participants to calculate the features of the lines, given that this time we could train the model on individual patterns for each participant and each movie trailer, rather than only the 27 fixed patterns (and their fixed features) used in the third-person studies. For the start value predictor, we integrated the first three seconds of the experience, as the first point is otherwise identical for all participants (i.e., it is the start point of the sliding scale) and participants might have different reaction times before they start moving the slider.

We utilized the ten-fold cross-validation approach to quantify how well each predictor predicted participant ratings (prediction accuracy = cross-validated Pearson's r). In line with hypothesis 2, participants' summaries of willingness to pay were significantly predicted by several features (table 4; figure 12). In this study, end value, integral, and slope were again

among the most predictive features, although sentiment score was the least predictive feature among all significant features, possibly because the movie genres made the words unreliable to interpret. For instance, someone might have enjoyed a horror trailer and yet described it as ‘terrifying’, which might not be a negative evaluation and yet be coded as having a negative sentiment.

TABLE 4

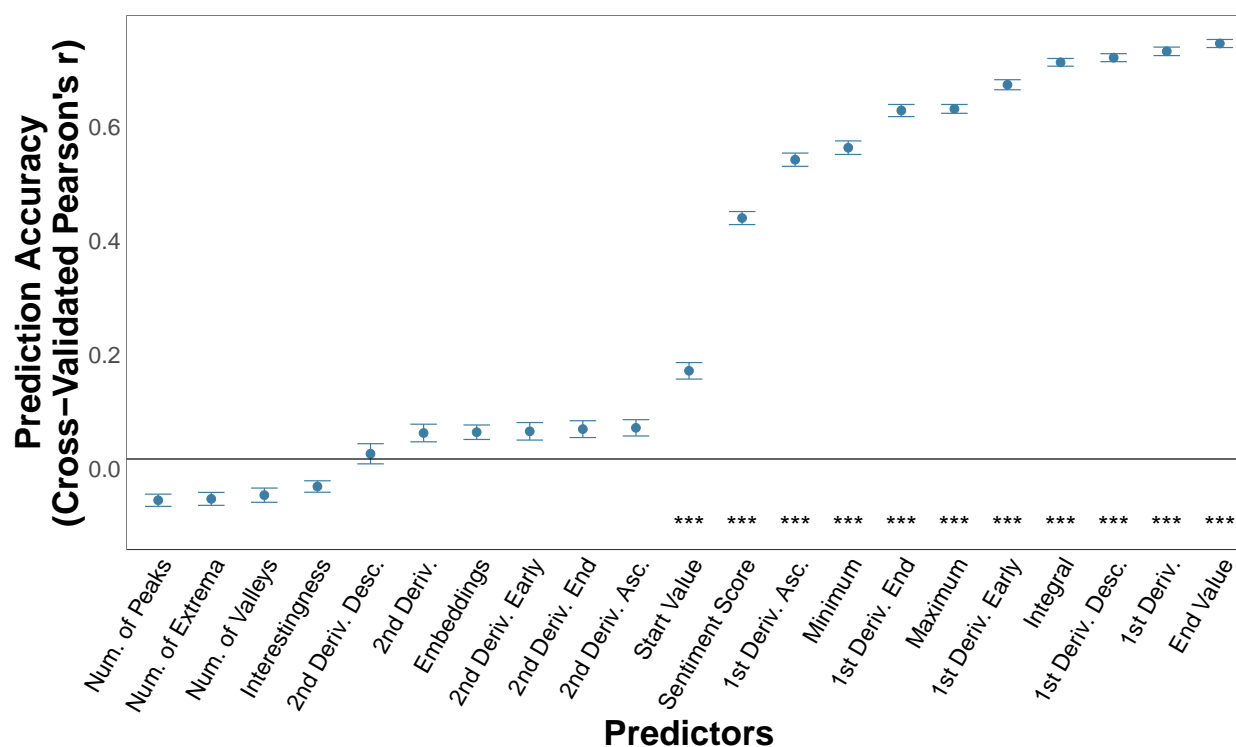
MEAN PERFORMANCE OF STATISTICALLY SIGNIFICANT PREDICTORS IN STUDY 4

Predictor	<i>r</i> Willingness to Buy
End value	0.75***
First derivative (unweighted)	0.73***
First derivative (desc. weights)	0.72***
Integral	0.71***
First derivative (early weights)	0.67***
Maximum value	0.63***
First derivative (end)	0.63***
Minimum value	0.56***
First derivative (asc. weights)	0.54***
Sentiment score	0.44***
Start value	0.17***

NOTE.— Significant predictors ranked best to worst. Predictors not featured were not statistically significant. *** = $p < .001$.

FIGURE 12

PERFORMANCE OF FEATURES IN PREDICTING WILLINGNESS TO PAY IN STUDY 4



NOTE.— Circles depict the mean prediction performance, and error bars depict 95% CI. Horizontal line and gray shadings depict mean and standard deviation of the prediction performance of a feature with randomly set values for each pattern. Stars beneath boxes indicate whether the root mean squared prediction errors are significantly less than that of a feature with random values. If the sample distribution was normal, we used a Welch two sample t-test, otherwise, we used a Wilcoxon signed-rank test. *** = $p < .001$. Lack of stars indicate non-significance.

To predict which movie participants wanted to enter the raffle for, we used a similar ten-fold cross validation approach as before, while using stratified sampling to retain the class proportions in each cross-validation fold (i.e., 1 out of 8 outcomes was a raffle choice).

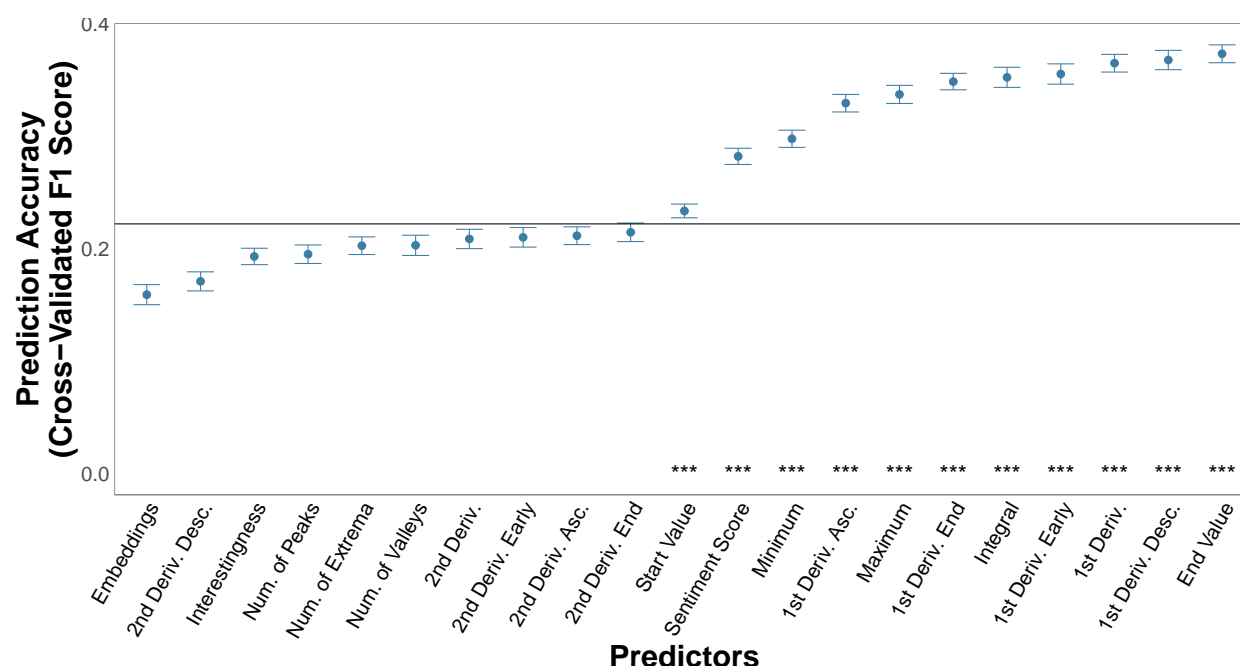
Since we were predicting the binary outcome of whether the participant chose the specific movie, we used the F1 Score to evaluate the model for each predictor. The F1 Score is a commonly used metric for imbalanced classification problems, since it is the harmonic mean (i.e., division of the number of items by the reciprocal of each item) of precision and recall. Precision shows the quality of positive predictions, i.e., the proportion of correctly predicted raffle choices divided by the total number of positive predictions. Recall shows the proportion of correctly predicted positives among all actual positives, i.e., the raffle choices. F1 Score tries to

equally value both metrics, thus avoiding the problem of scoring the model highly for taking the ‘lazy’ strategy of just predicting all zeros or ones. Because we have an imbalanced classification problem (i.e., there is a single picked movie vs seven unpicked ones for each participant), we utilized a weighted logistic regression model to emphasize predicting the chosen movie, and used the inverse of the class ratio for the weights (Brownlee 2020).

Of the participants who provided no more than one maximum willingness to pay rating across trailers (56 provided more than one), 74% of participants picked as their raffle choice the movie for which they were most willing to pay, suggesting that it would be fairly challenging to predict their raffle choices based on features of their experiences. Given this upper bound on how well we expected to perform, we found that the F1 Score for the predictors performed reasonably, i.e., the highest F1 Score among each predictor approached 0.4. We also noticed that the same significant features predicting willingness to pay (figure 12) were significant for predicting the raffle choice (figure 13), and that ordering them the same and correlating their outcomes yielded a high correlation ($r = .97, p < .001$).

FIGURE 13

PERFORMANCE OF FEATURES IN PREDICTING RAFFLE CHOICE IN STUDY 4



NOTE.— Circles depict the mean prediction performance, and error bars depict 95% CI. Stars beneath boxes indicate whether the distribution of correlations is significantly different from 0.222 (the horizontal line), which depicts an F1 Score for predicting all ones (the 'lazy' strategy with the highest F1 Score, reflecting no meaningful learning). We used a Welch two sample t-test if the sample distribution was normal, otherwise we used a Wilcoxon signed-rank test. *** = $p < .001$. A lack of stars indicates non-significance.

Discussion

In line with hypothesis 1, we intuitively observed that willingness to pay ratings were most positive for patterns that improved, followed by fluctuations, then by deteriorations. Further, in line with hypothesis 2, these judgments were predicted by several features of the pattern rather than just one, including (from least to most predictive): sentiment score, minimum, maximum, start value, integral, first derivative, and (the best predictor of all) end value. Likewise, we found that end value was the best predictor of raffle choice, underscoring its influence on both judgments and behavior.

Finally, we tested whether these first-person results correlated with the third-person results of the previous studies. First, we had to assign the 27 clusters of first-person patterns to

the closest corresponding third-person patterns from the previous studies. To do so, we calculated the sum of absolute differences between each pattern of the first-person cluster and each pattern used in third person studies. Based on these mean-squared errors, we assigned, using the Hungarian algorithm (Kuhn 1955), each cluster to a third-person pattern such that it had the minimum overall error (figure S8). Supporting hypothesis 5, the previous effects generalized reasonably well to first-person experiences (table 5), suggesting that a similar psychological process is at play when summarizing first and third-person experiences.

TABLE 5
CORRELATIONS BETWEEN RATINGS ACROSS STUDIES 1, 3, S2 AND 4

Customer journey satisfaction v. movie willingness to buy	0.61***
Customer journey personal desirability v. movie willingness to buy	0.55**
Hiring v. movie willingness to buy	0.63***
Life meaningfulness v. movie willingness to buy	0.77***
Life personal desirability v. movie willingness to buy	0.73***

GENERAL DISCUSSION

How do consumers summarize continuous experiences like watching a trailer or undergoing an entire customer journey with a firm? Our studies investigated this question by both generating and capturing a wide variety of experience patterns, and quantifying a comprehensive set of features of those patterns. Helping to adjudicate between whether smooth or fluctuating journeys are most effective (Siebert et al. 2020) at the level of the mental experience itself, we found that consumers were most delighted by experience patterns that were

consistently positive or improving, followed by fluctuating patterns, and then by ones that were consistently negative or deteriorating.

Across all studies, we found that similar literal and conceptual features—such as start value, integral, maximum and minimum, sentiment, slope, and especially the end value—performed above chance at predicting consumer judgments and behavior. In particular, the end value was consistently one of the top two best predictors, underscoring the importance of ending customer experiences on a high note. Also, in all experience domains except for watching movie trailers, we found that a sentiment score of the language participants used to describe the experience was the very best predictor. This indicates that, when asked to describe an experience, consumers spontaneously represented it in evaluative terms, i.e., whether the experience was going well or poorly overall. The exception was in the study on movie trailers, where descriptions of the movie genre and content (e.g., “terrifying”, “romantic”) likely added noise to the sentiment scores; although sentiment score still significantly predicted the outcome variables.

Looking at the non-significant predictors, we found that acceleration was not a significant predictor in any study, indicating that summarization does not factor in acceleration as much as it does the velocity and end point of an experience, i.e., whether the experience improves/deteriorates and ends on a high note. Further, the total extrema feature was not predictive in all studies, perhaps because both successful and unsuccessful patterns can have a similar number of peaks and valleys. For example, both a linearly increasing and a linearly decreasing pattern have no peaks nor valleys, yet the increasing function is perceived to be satisfying whereas the decreasing one is not.

There was also some interesting variation in the ordering of features depending on the domain, which might have been due to the different temporal scales and perceived degree of control between those domains. For relatively shorter experiences like customer journeys, interview performances and movie trailers, the first derivative tended to be more important than the integral, whereas for the longer experience of a life the integral was more important than the first derivative. Perhaps this is because the first derivative (aka slope) seems less controllable against the backdrop of an entire life than the sum of one's happiness. Despite these differences, we found that consumers largely summarized patterned experiences from different perspectives similarly. The similarity across perspectives also speaks against the idea that the only reason endings are important in summarization is that they already reflect partial evaluations of the ongoing experience thus far, or because they indicate the effect of more recent experiences in memory on evaluations (aka recency effects in memory). While such explanations are possible for first-person experiences, they do not apply for retrospective evaluations of third-person experiences.

We also found evidence that consumers are unaware of the psychological processes underlying their summaries (study 2), in line with work suggesting that some mental modules operate automatically without consumers being introspectively aware of their psychological processes (Kahneman 2011; Logan and Cowan 1984). If anything, participants tended to intuitively place more importance on predictors that were less important in practice. The automaticity of summarization suggests that marketers can influence it without consumers realizing how they did this.

Theoretical Contributions

This work makes theoretical contributions to work on customer journeys, summarization, and predicting the success of content.

Regarding customer journeys, the findings extend beyond stage models of the customer journey (Bettman and Park 1980; Demmers et al. 2020; Grewal and Roggeveen 2020; Hamilton et al. 2021; Schamp et al. 2019). While stage models are useful conceptualizations of customer journeys, the present research suggests that it is also in a firm's best interests to collect higher resolution data into the continuous pattern of the customer journey. Further, the findings help arbitrate among theoretical positions on what types of experience patterns are most satisfying. One view is that customer journeys should be smooth, while another is that they should constantly fluctuate because this is more exciting (Siebert et al. 2020). Supporting the smooth view, we found that consistently positive experiences were judged as more satisfying than fluctuating ones, at least at the level of the mental experience itself. At the same time, it is not that consumers were averse to any fluctuation whatsoever, since they rated almost just as positively experiences which ended pleasantly, constantly improved, or had a distinct peak; what these fluctuations have in common is that they improve over time and/or end on a high note. Interestingly, participants also judged that fluctuating experiences could be more satisfying than they were personally desirable, perhaps because they wanted to avoid unpredictability, effort, and adversity. With that said, satisfaction and desirability converged for trajectories that were viewed as either very satisfying or unsatisfying.

The findings also extend our understanding of summarization. In line with previous work, we found that the peak and end values of a trajectory predict how the trajectory is summarized (Diener et al. 2001; Fredrickson and Kahneman 1993; Newman et al. 2010; O'Brien and

Ellsworth 2012; Redelmeier and Kahneman 1996). At the same time, our more exhaustive set of experience patterns and features revealed that peaks and ends were just one piece of the puzzle—first derivatives, integrals and language sentiment mattered too. Moreover, the most consistently top performing feature across studies was the endpoint, underscoring the importance of endings across a wide array of experience patterns than explored in previous work and while pitting this feature against a more extensive set of contending features. Future work can also explore whether the current results converge with those of less theory-driven approaches, such as functional regression (Hui et al. 2014).

Relatedly, we contribute to work on predicting the success of content such as stories, academic articles, social media posts, songs, and movies (Berger and Packard 2018; Laurino Dos Santos and Berger 2022; Reagan et al. 2016; Toubia et al. 2021). Some previous work has treated features of the content itself as a *proxy* of mental states, e.g., treating fluctuating content sentiment as a proxy for consumer engagement levels (Berger et al. 2021). We find that, even if fluctuations are desirable at the level of the content itself, they are undesirable at the level of the mental experience (as compared to consistently positive or improving experiences). More broadly, one way to reconcile the current findings with previous work on both the success of content (Berger et al. 2021) and the importance of fluctuating customer journeys (Siebert et al. 2020) is that perhaps fluctuating *content* along the customer journey creates consistently positive or improving *mental states* for the consumers of those journeys.

Practical Implications

Our findings suggest that, beyond striving for consistently positive customer experiences, managers can invest their limited resources in creating experiences that improve over time and have a climatic end and peak point. First, they can invest in later touchpoints along the journey, and especially endings, as when Disney World offers fireworks at the end of a day, or when furniture and retail firm Crate & Barrel conveniently brings all packages to the customer's car. Second, given a set of stronger and weaker offerings, managers can ramp them up from least to most impressive, such as increasingly entertaining events on a cruise, or increasingly high stakes matches in the FIFA World Cup (e.g., quarter, semi, and final matches). Third, they can invest in making one touchpoint particularly impressive, to create a memorable peak, as when patrons to Santouka Ramen are greeted by a chorus of servers, or when Yoga studios offer 'hands on' scented treatment during the session.

In lieu of gaining access to visibility into the entire customers journey, which may be costly or perceived as invasive by customers, a viable alternative is to use a sentiment score of the language which customers use to summarize their experiences. Thus, managers may want to proactively collect such data as a window into the how their customer journeys are being spontaneously represented in consumer's minds, as via quick prompts, customer feedback surveys, and reviews.

REFERENCES

- Aaker, David A, Douglas M Stayman, and Michael R Hagerty (1986), "Warmth in Advertising: Measurement, Impact, and Sequence Effects," *Journal of Consumer Research*, 12 (4), 365-81.
- Alter, Adam (2017), *Irresistible: Why We Can't Stop Checking, Scrolling, Clicking and Watching*: Random House.
- Ariely, Dan (1998), "Combining Experiences over Time: The Effects of Duration, Intensity Changes and on-Line Measurements on Retrospective Pain Evaluations," *Journal of Behavioral Decision Making*, 11 (1), 19-45.

- Bargh, John A (2002), "Losing Consciousness: Automatic Influences on Consumer Judgment, Behavior, and Motivation," *Journal of Consumer Research*, 29 (2), 280-85.
- Bargh, John A and Tanya L Chartrand (1999a), "The Unbearable Automaticity of Being," *American Psychologist*, 54 (7), 462-79.
- (1999b), "The Unbearable Automaticity of Being," *American Psychologist*, 54 (7), 462.
- Baumgartner, Hans, Mita Sujan, and Dan Padgett (1997), "Patterns of Affective Reactions to Advertisements: The Integration of Moment-to-Moment Responses into Overall Judgments," *Journal of Marketing Research*, 34 (2), 219-32.
- Berger, Jonah, Yoon Duk Kim, and Robert Meyer (2021), "What Makes Content Engaging? How Emotional Dynamics Shape Success," *Journal of Consumer Research*, 48 (2), 235-50.
- Berger, Jonah and Grant Packard (2018), "Are Atypical Things More Popular?," *Psychological Science*, 29 (7), 1178-84.
- Bettman, James R and C Whan Park (1980), "Effects of Prior Knowledge and Experience and Phase of the Choice Process on Consumer Decision Processes: A Protocol Analysis," *Journal of Consumer Research*, 7 (3), 234-48.
- Bhargave, Rajesh and Nicole Votolato Montgomery (2013), "The Social Context of Temporal Sequences: Why First Impressions Shape Shared Experiences," *Journal of Consumer Research*, 40 (3), 501-17.
- Blei, David M, Andrew Y Ng, and Michael I Jordan (2003), "Latent Dirichlet Allocation," *Journal of Machine Learning Research*, 3, 993-1022.
- Brownlee, Jason (2020), *Imbalanced Classification with Python: Better Metrics, Balance Skewed Classes, Cost-Sensitive Learning: Machine Learning Mastery*.
- Carmon, Ziv and Daniel Kahneman (1996), "The Experienced Utility of Queuing: Real Time Affect and Retrospective Evaluations of Simulated Queues," *Duke University: Durham, NC, USA*.
- Cer, Daniel, Yinfei Yang, Sheng-yi Kong, Nan Hua, Nicole Limtiaco, Rhomni St John, Noah Constant, Mario Guajardo-Cespedes, Steve Yuan, and Chris Tar (2018), "Universal Sentence Encoder," *arXiv preprint arXiv:1803.11175*.
- De Freitas, Julian and George A Alvarez (2018), "Your Visual System Provides All the Information You Need to Make Moral Judgments About Generic Visual Events," *Cognition*, 178, 133-46.
- De Freitas, Julian, Mina Cikara, Igor Grossmann, and Rebecca Schlegel (2018), "Moral Goodness Is the Essence of Personal Identity," *Trends in Cognitive Sciences*, 22 (9), 739-40.
- De Freitas, Julian and Alon Hafri (2022), "Moral Thin-Slicing," *Harvard Business School Marketing Unit Working Paper* (23-002).
- De Freitas, Julian, Zarema Khon, Pechthida Kim, and Samuel Johnson (2022), "Moral Deteriorations Sever Firm Identity," *Harvard Business School Marketing Unit Working Paper* (22-077).
- Deighton, John (1992), "The Consumption of Performance," *Journal of Consumer Research*, 19 (3), 362-72.
- Deighton, John, Daniel Romer, and Josh McQueen (1989), "Using Drama to Persuade," *Journal of Consumer Research*, 16 (3), 335-43.

- Demmers, Joris, Jesse WJ Weltevreden, and Willemijn M van Dolen (2020), "Consumer Engagement with Brand Posts on Social Media in Consecutive Stages of the Customer Journey," *International Journal of Electronic Commerce*, 24 (1), 53-77.
- Dhillon, Paramveer S and Sinan Aral (2021), "Modeling Dynamic User Interests: A Neural Matrix Factorization Approach," *Marketing Science*, 40 (6), 1059-80.
- Diener, Ed, Derrick Wirtz, and Shigehiro Oishi (2001), "End Effects of Rated Life Quality: The James Dean Effect," *Psychological science*, 12 (2), 124-28.
- Edelman, David C and Marc Singer (2015), "Competing on Customer Journeys," *Harvard Business Review*, 93 (11), 88-100.
- Eyal, Nir (2014), *Hooked: How to Build Habit-Forming Products*: Penguin.
- Fazio, Russell H, David M Sanbonmatsu, Martha C Powell, and Frank R Kardes (1986), "On the Automatic Activation of Attitudes," *Journal of Personality and Social Psychology*, 50 (2), 229.
- Firestone, Chaz and Brian J Scholl (2016), "Cognition Does Not Affect Perception: Evaluating the Evidence for "Top-Down" Effects," *Behavioral and Brain Sciences*, 39.
- Fredrickson, Barbara L and Daniel Kahneman (1993), "Duration Neglect in Retrospective Evaluations of Affective Episodes," *Journal of Personality and Social Psychology*, 65 (1), 45-55.
- Frow, Pennie and Adrian Payne (2007), "Towards the 'Perfect' customer Experience," *Journal of Brand Management*, 15 (2), 89-101.
- Garnefeld, Ina and Lena Steinhoff (2013), "Primacy Versus Recency Effects in Extended Service Encounters," *Journal of Service Management*.
- Grewal, Dhruv and Anne L Roggeveen (2020), "Understanding Retail Experiences and Customer Journey Management," *Journal of Retailing*, 96 (1), 3-8.
- Hamilton, Ryan, Rosellina Ferraro, Kelly L Haws, and Anirban Mukhopadhyay (2021), "Traveling with Companions: The Social Customer Journey," *Journal of Marketing*, 85 (1), 68-92.
- Hansen, David E and Peter J Danaher (1999), "Inconsistent Performance During the Service Encounter: What's a Good Start Worth?," *Journal of Service Research*, 1 (3), 227-35.
- Harman, David M and Michael C Porter (2021), "Your Receipt Is in the Bag: Service and Temporal Effects as Factors of Customer Engagement Formation During Acquisition," *Journal of Retailing and Consumer Services*, 62, 102603.
- Hastie, Trevor, Robert Tibshirani, Jerome H Friedman, and Jerome H Friedman (2009), *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, Vol. 2: Springer.
- Hildebrand, Christian and Tobias Schlager (2019), "Focusing on Others before You Shop: Exposure to Facebook Promotes Conventional Product Configurations," *Journal of the Academy of Marketing Science*, 47 (2), 291-307.
- Homburg, Christian, Danijel Jozić, and Christina Kuehnl (2017), "Customer Experience Management: Toward Implementing an Evolving Marketing Concept," *Journal of the Academy of Marketing Science*, 45 (3), 377-401.
- Hsee, Christopher K and Robert P Abelson (1991), "Velocity Relation: Satisfaction as a Function of the First Derivative of Outcome over Time," *Journal of Personality and Social Psychology*, 60 (3), 341.

- Hsee, Christopher K, Peter Salovey, and Robert P Abelson (1994), "The Quasi-Acceleration Relation: Satisfaction as a Function of the Change of Velocity of Outcome over Time," *Journal of Experimental Social Psychology*, 30 (1), 96-111.
- Hughes, G David (1992), "Realtime Response Measures Redefine Advertising Wearout," *Journal of Advertising Research*, 32 (3), 61-77.
- Hui, Sam K, Tom Meyvis, and Henry Assael (2014), "Analyzing Moment-to-Moment Data Using a Bayesian Functional Linear Model: Application to Tv Show Pilot Testing," *Marketing Science*, 33 (2), 222-40.
- Hung, Iris W and Anirban Mukhopadhyay (2012), "Lenses of the Heart: How Actors' and Observers' Perspectives Influence Emotional Experiences," *Journal of Consumer Research*, 38 (6), 1103-15.
- Kahneman, Daniel (2011), *Thinking, Fast and Slow*: Macmillan.
- Kahneman, Daniel, Barbara L Fredrickson, Charles A Schreiber, and Donald A Redelmeier (1993), "When More Pain Is Preferred to Less: Adding a Better End," *Psychological Science*, 4 (6), 401-05.
- Kahn, Barbara, Rebecca Ratner, and Daniel Kahneman (1997), "Patterns of Hedonic Consumption over Time," *Marketing Letters*, 8 (1), 85-96.
- Kausel, Edgar E, Satoris S Culbertson, and Hector P Madrid (2016), "Overconfidence in Personnel Selection: When and Why Unstructured Interview Information Can Hurt Hiring Decisions," *Organizational Behavior and Human Decision Processes*, 137, 27-44.
- Knobe, Joshua, Sandeep Prasada, and George E Newman (2013), "Dual Character Concepts and the Normative Dimension of Conceptual Representation," *Cognition*, 127 (2), 242-57.
- Kohavi, Ron (1995), "A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection," in *Ijcai*, Vol. 14: Montreal, Canada, 1137-45.
- Kuehnl, Christina, Danijel Jozic, and Christian Homburg (2019), "Effective Customer Journey Design: Consumers' Conception, Measurement, and Consequences," *Journal of the Academy of Marketing Science*, 47 (3), 551-68.
- Kuhn, Harold W (1955), "The Hungarian Method for the Assignment Problem," *Naval Research Logistics Quarterly*, 2 (1-2), 83-97.
- Laurino Dos Santos, Henrique and Jonah Berger (2022), "The Speed of Stories: Semantic Progression and Narrative Success," *Journal of Experimental Psychology: General*.
- Legg, Angela M and Kate Sweeny (2014), "Do You Want the Good News or the Bad News First? The Nature and Consequences of News Order Preferences," *Personality and Social Psychology Bulletin*, 40 (3), 279-88.
- Lemon, Katherine N and Peter C Verhoef (2016), "Understanding Customer Experience Throughout the Customer Journey," *Journal of Marketing*, 80 (6), 69-96.
- Lin, Yan, Dai Yao, and Xingyu Chen (2021), "Happiness Begets Money: Emotion and Engagement in Live Streaming," *Journal of Marketing Research*, 58 (3), 417-38.
- Loewenstein, George F and Dražen Prelec (1993), "Preferences for Sequences of Outcomes," *Psychological review*, 100 (1), 91.
- Logan, Gordon D and William B Cowan (1984), "On the Ability to Inhibit Thought and Action: A Theory of an Act of Control," *Psychological Review*, 91 (3), 295.
- MacQueen, J (1967), "Classification and Analysis of Multivariate Observations," in *5th Berkeley Symposium on Mathematical Statistics and Probability*, 281-97.

- Madrigal, Robert and Colleen Bee (2005), "Suspense as an Experience of Mixed Emotions: Feelings of Hope and Fear While Watching Suspenseful Commercials," *ACR North American Advances*.
- Nelson, Leif D, Tom Meyvis, and Jeff Galak (2009), "Enhancing the Television-Viewing Experience through Commercial Interruptions," *Journal of Consumer Research*, 36 (2), 160-72.
- Newman, George E, Kristi L Lockhart, and Frank C Keil (2010), "'End-of-Life' Biases in Moral Evaluations of Others," *Cognition*, 115 (2), 343-49.
- O'Brien, Ed and Phoebe C Ellsworth (2012), "Saving the Last for Best: A Positivity Bias for End Experiences," *Psychological science*, 23 (2), 163-65.
- Olian, Judy D, Donald P Schwab, and Yitchak Haberfeld (1988), "The Impact of Applicant Gender Compared to Qualifications on Hiring Recommendations: A Meta-Analysis of Experimental Studies," *Organizational Behavior and Human Decision Processes*, 41 (2), 180-95.
- Osgood, Charles E (1952), "The Nature and Measurement of Meaning," *Psychological Bulletin*, 49 (3), 197-237.
- Pham, Michel Tuan, Joel B Cohen, John W Pracejus, and G David Hughes (2001), "Affect Monitoring and the Primacy of Feelings in Judgment," *Journal of Consumer Research*, 28 (2), 167-88.
- Pihur, Vasyi, Susmita Datta, and Somnath Datta (2007), "Weighted Rank Aggregation of Cluster Validation Measures: A Monte Carlo Cross-Entropy Approach," *Bioinformatics*, 23 (13), 1607-15.
- Ramanathan, Suresh and Ann L McGill (2007), "Consuming with Others: Social Influences on Moment-to-Moment and Retrospective Evaluations of an Experience," *Journal of Consumer Research*, 34 (4), 506-24.
- Ratner, Rebecca K, Barbara E Kahn, and Daniel Kahneman (1999), "Choosing Less-Preferred Experiences for the Sake of Variety," *Journal of Consumer Research*, 26 (1), 1-15.
- Reagan, Andrew J, Lewis Mitchell, Dilan Kiley, Christopher M Danforth, and Peter Sheridan Dodds (2016), "The Emotional Arcs of Stories Are Dominated by Six Basic Shapes," *EPJ Data Science*, 5 (1), 1-12.
- Redelmeier, Donald A and Daniel Kahneman (1996), "Patients' Memories of Painful Medical Treatments: Real-Time and Retrospective Evaluations of Two Minimally Invasive Procedures," *Pain*, 66 (1), 3-8.
- Ross, William T and Itamar Simonson (1991), "Evaluations of Pairs of Experiences: A Preference for Happy Endings," *Journal of Behavioral Decision Making*, 4 (4), 273-82.
- Schamp, Christina, Mark Heitmann, and Robin Katzenstein (2019), "Consideration of Ethical Attributes Along the Consumer Decision-Making Journey," *Journal of the Academy of Marketing Science*, 47 (2), 328-48.
- Schmitt, Bernd (1999), "Experiential Marketing," *Journal of Marketing Management*, 15 (1-3), 53-67.
- Schweidel, David A, Yakov Bart, J Jeffrey Inman, Andrew T Stephen, Barak Libai, Michelle Andrews, Ana Babić Rosario, Inyoung Chae, Zoey Chen, and Daniella Kupor (2022), "How Consumer Digital Signals Are Reshaping the Customer Journey," *Journal of the Academy of Marketing Science*, 50, 1257-76.

- Sharps, Daron L and Cameron Anderson (2021), "Social Class Background, Disjoint Agency, and Hiring Decisions," *Organizational Behavior and Human Decision Processes*, 167, 129-43.
- Shtudiner, Ze'ev (2019), "The Impact of Attractiveness on Retail Managers' Decisions: Labor Market Discrimination When Hiring Salespeople," *Available at SSRN 3479010*.
- Siebert, Anton, Ahir Gopaldas, Andrew Lindridge, and Cláudia Simões (2020), "Customer Experience Journeys: Loyalty Loops Versus Involvement Spirals," *Journal of Marketing*, 84 (4), 45-66.
- Teixeira, Thales, Michel Wedel, and Rik Pieters (2012), "Emotion-Induced Engagement in Internet Video Advertisements," *Journal of Marketing Research*, 49 (2), 144-59.
- Thorson, Esther and Marian Friestad (1989), "The Effects of Emotion on Episodic Memory for Television Commercials," in *Cognitive and Affective Responses to Advertising*, ed. Cafferata P and Tybout A: Lexington Books, 305-25.
- Todorov, Alexander, Anesu N Mandisodza, Amir Goren, and Crystal C Hall (2005), "Inferences of Competence from Faces Predict Election Outcomes," *Science*, 308 (5728), 1623-26.
- Toubia, Olivier, Jonah Berger, and Jehoshua Eliashberg (2021), "How Quantifying the Shape of Stories Predicts Their Success," *Proceedings of the National Academy of Sciences*, 118 (26), e2011695118.
- Tsay, Chia-Jung (2013), "Sight over Sound in the Judgment of Music Performance," *Proceedings of the National Academy of Sciences*, 110 (36), 14580-85.
- Vanden Abeele, Piet and Douglas L MacLachlan (1994), "Process Tracing of Emotional Responses to Tv Ads: Revisiting the Warmth Monitor," *Journal of Consumer Research*, 20 (4), 586-600.
- Verhoef, Peter C, Gerrit Antonides, and Arnoud N De Hoog (2004), "Service Encounters as a Sequence of Events: The Importance of Peak Experiences," *Journal of Service Research*, 7 (1), 53-64.
- Vonnegut, K (1995), "Shapes of Stories," <https://www.youtube.com/watch?v=oP3c1h8v2ZQ>.
- Winet, Yuji K and Ed O'Brien (2022), "Ending on a Familiar Note: Perceived Endings Motivate Repeat Consumption," *Journal of Personality and Social Psychology*.
- Woltman Elpers, Josephine LCM, Ashesh Mukherjee, and Wayne D Hoyer (2004), "Humor in Television Advertising: A Moment-to-Moment Analysis," *Journal of Consumer Research*, 31 (3), 592-98.
- Woltman Elpers, Josephine LCM, Michel Wedel, and Rik GM Pieters (2003), "Why Do Consumers Stop Viewing Television Commercials? Two Experiments on the Influence of Moment-to-Moment Entertainment and Information Value," *Journal of Marketing Research*, 40 (4), 437-53.
- Zhang, Qiang, Wenbo Wang, and Yuxin Chen (2020), "Frontiers: In-Consumption Social Listening with Moment-to-Moment Unstructured Data: The Case of Movie Appreciation and Live Comments," *Marketing Science*, 39 (2), 285-95.