Journal of Economic Behavior and Organization 224 (2024) 771–784

Contents lists available at ScienceDirect

Journal of Economic Behavior and Organization

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

Research paper

And suddenly, the rain! When surprises shape experienced utility

Paolo Figini [a,b], Veronica Leoni [b,c,*], Laura Vici [b,d]

[a] *Department of Economics, University of Bologna, Italy*
[b] *Center for Advanced Studies in Tourism, University of Bologna, Italy*
[c] *Department of Applied Economics, University of the Balearic Islands, Spain*
[d] *Department of Statistical Sciences 'Paolo Fortunati' University of Bologna, Italy*

ARTICLE INFO

JEL classification:
D83
D91
L81
Q54

Keywords:
Surprise
Weather bias
Experienced utility
Online reviews
Rating platforms

ABSTRACT

This study examines how unexpected exogenous events, labelled as *suprises*, affect the utility of experience goods reported in online rating systems. Using over 300,000 reviews of accommodation services listed on Booking.com, the research investigates whether online ratings capture the impact of surprises related to meteorological conditions and whether they create additional biases in service evaluation. The study finds that sudden changes in weather conditions have a significant impact on experienced utility, with the effect varying based on the direction of the surprise. Additionally, in line with the hedonic adaption theory, we find that the duration of consumption moderates the surprise effect, reducing its impact on reported utility.

1. Introduction

When purchasing products and services, consumers often base their choice on explicit quality signals, which may come from several sources, such as brand reputation, public certifications, or user-generated content. The greater the asymmetry of information between the seller and prospective buyers, the higher the need to obtain timely and reliable metrics of quality. This is especially relevant for markets with no well-defined quality labels, for first-time buyers, and when dealing with experience goods, whose quality can only be assessed during or after consumption (Nelson, 1970). While some commodities are cheap, and their purchase does not entail a burdensome decision process, buying more expensive commodities involves carefully acquiring information to form reliable expectations. Thanks to the steady progress of online platforms and social media, it is now extremely common for subjects to access first-hand and spontaneous feedback from previous customers on their consumption experience and their degree of satisfaction with goods and services. Well-known examples include TripAdvisor for restaurant meals, Booking.com for accommodation services, and Amazon or TrustPilot for commodities. The popularity of online review systems is reflected in an emerging and growing literature with both theoretical and applied contributions on the reviewing mechanism of different platforms (Ifrach et al., 2019; Helmers et al., 2019; Greiff and Paetzel, 2020; Reimers and Waldfogel, 2021; Mayzlin et al., 2014; Acemoglu et al., 2022).

As discussed in Acemoglu et al. (2022), when a subject decides to post a review after consuming a product, their score reflects what is termed as *material utility*. This indicates the extent to which the good or service met prior expectations. However, existing literature has raised concerns about various factors potentially biasing such post-purchase evaluations. Among the most discussed are herding behaviours since current raters' evaluations are often shaped by prior ratings (Chevalier and Mayzlin, 2006; Cicognani et al., 2021; DeMarzo et al., 2003; Manski, 1993; Sunder et al., 2019). Another factor is the existence of selection effects, which

* Correspondence to: Department of Applied Economics, University of the Balearic Islands, Carretera de Valldemossa km 7.5, 07122, Palma de Mallorca, Spain.
E-mail address: veronica.leoni@uib.es (V. Leoni).

https://doi.org/10.1016/j.jebo.2024.06.026
Received 6 June 2023; Received in revised form 17 May 2024; Accepted 18 June 2024
Available online 6 July 2024
0167-2681/

Table 1
Classification of factors affecting the review.

		Anticipation	
		Expected	*Unexpected (surprise)*
Dimension	*Intrinsic service quality (related factors)*	(1) Fundamental service	(2) Fundamental shock
	Situational factors (unrelated factors)	(3) Non-fundamental service	(4) Non-Fundamental shock

lead to a J-shaped distribution of reviews (Hu et al., 2017; Li and Hitt, 2008), or brag-or-moan effect, with only consumers who are highly satisfied or dissatisfied choosing to share quality information (Schoenmueller et al., 2020).

In the case of services, the situation becomes more complex, as the quality of the experience depends not only on the intrinsic service quality or the meeting of prior expectations but also on situational factors and exogenous shock, which are outside of producers' control. This is critical, as it may cast doubt on the ability of online reviews to precisely reveal the quality of services and, hence, diminish their informational value for prospective buyers who rely on such information to make consumption decisions. A classic example of an experience good that aligns with the description provided above is found in the travel industry and is represented by accommodation services. Here, a typical external factor is weather conditions, which can significantly impact the enjoyment of the overall travel experience (Brandes and Dover, 2022). On the one hand, it can be argued that weather information is already incorporated into prior beliefs and should not shock the reported utility since historical climate conditions and weather forecasts are public information available at virtually no cost. On the other hand, a potentially non-neutral impact could arise from unexpected weather events, i.e., those not properly anticipated and, hence, not part of the consumer's set of information available before the experience consumption. In this context, the mismatch between forecasts and realized weather brings to what can be defined as a *surprise* for the consumer.

A growing body of literature has focused on the role of surprises in shaping utility (Ely et al., 2015; Bizzozero et al., 2016), but from a partially different perspective. The aforementioned works examine the dynamics of surprisingness and investigate the satisfaction derived from specific suspense and surprise interactions. Notably, Ely et al. (2015) is set within a principal–agent framework, where the principal selects a degree of information disclosure to generate a flow of suspense and surprise, with the goal of maximizing consumer utility and, consequently, enjoyment. This literature is primarily related to media and sports contexts.

In contrast, our research focuses on the role of surprises when evaluating services and where the surprise is considered an exogenous shock rather than an intrinsic aspect of the service being assessed. In this context, a surprising event represents a discrepancy between the expectation and the realization of a particular event (Baccan et al., 2015; Schützwohl, 1998; Derbaix and Vanhamme, 2003). By adopting this unique perspective, we aim to investigate if and to what extent the experienced utility reported by online rating platforms captures such exogenous shocks and, in doing so, unravel a potentially additional distortion of rating systems.

To better clarify the issue, let us represent in Table 1 the elements that likely affect the service rating. On the one hand, factors can be classified according to whether they are directly linked to the quality of the service to be reviewed (intrinsic factors) or not (situational factors). On the other hand, factors can be classified according to whether they can be anticipated (and therefore be part of the expectations) or not. These last factors, that arrive unexpectedly, are the surprises. Overall, the double-entry table highlights four types of factors.

Theory tells us that reviews should focus on factors of type (1), the intrinsic quality of the service (e.g., in the accommodation industry, if the room with a sea view that the customer has booked conforms to their expectations) and type (2), how 'fundamental shocks' affect this expected quality (e.g., the review might be negatively be affected if the hotel is under renovation and has scaffolding, or positively affected if the customer gets the suite as a free upgrade). There is also relevant literature in tourism (Gössling et al., 2016; Jeuring, 2017; Jeuring and Peters, 2013; Brandes and Dover, 2022) and other industries (Connolly, 2008; Lee et al., 2014; Huysmans, 2002) showing that external events and situational factors of type (3) have an influence (and the weather is certainly relevant for the tourism experience, although not directly related to the quality of the accommodation service).

Our contribution to this literature is that also shocks of type (4) on situational elements (non-fundamental shocks) are important when the experience utility is reported. Indeed, we demonstrate that surprises related to exogenous events, besides the exogenous events *per se*, are captured by online ratings. We argue that type (4) under investigation in this paper constitutes a bias because it affects the information content of the review without being relevant information to future customers.[1] To summarize, what we

[1] It is important to stress the difference between the impact played by an external factor and a related surprise. Let us consider, to clarify the issue, the impact of rain on the experience of eating in a Venice restaurant. On a rainy day, customers are obliged to eat inside. According to the literature and intuition, the rating of the customer eating inside on a rainy day should, *ceteris paribus*, be lower than that of a customer eating outside on a sunny day because of the lower utility associated with this experience. We claim that customers who book a table in the restaurant (or enter the restaurant) on a rainy day already expect (and therefore internalize) the negative impact of such a state of the world. What they cannot internalize is the sudden change in the conditions. Therefore, given the realization of the rainy state of the world, reviews should be worse for those customers who book with expected sunny weather to eat outside, and then, because of a sudden worsening in meteorological conditions, they are compelled to move inside. Or, vice versa, given the expectation of rain (and hence of eating inside), reviews should be better for those who are moved outside on the terrace because the sun suddenly shines.

specifically test in this work is not the impact of the weather *per se* but how the disconfirmation of priors and expectations related to the weather drives the rating score.

To this end, we analyse more than 300,000 online reviews for accommodation providers listed on Booking.com, the leading search engine for lodging services, in two popular Italian destinations (Milan and Venice) between September 2019 and February 2020. The period under investigation is chosen for consistency reasons, allowing us to analyse only ratings before the COVID-19 outbreak and after the reform of the Booking.com reviewing system introduced by the platform in September 2019, constituting a structural break (Leoni and Boto-García, 2023). With regard to the surprises, we look at the potential mismatch between the weather forecasts available days before travelling and the realized weather when on holiday. The empirical strategy aims to disclose this 'surprise' effect on individual ratings, conditional on the reviewers' characteristics, providers' time-invariant features, prices, and weather conditions.

Given the specific type of surprise under consideration in our study – unexpected changes in weather conditions as compared to publicly available forecasts – our research also adds to the ongoing discussion in the literature about the influence of public information on individual and organizational behaviour (Figini et al., 2022; Angelini et al., 2023). Therefore, our work not only assesses the existence of an additional bias in service evaluations provided by online platforms but also has broader implications for the role of public information in decision-making processes.

Our findings provide conclusive evidence of a statistically significant impact of weather surprises on the reported experienced utility. Moreover, we uncover an interesting moderating role played by the consumption span, which, we argue, mitigates the effects of the weather mismatch, a finding that aligns with the hedonic adaptation theory (Frederick and Loewenstein, 1999). Results are robust to several specifications and, differently from most of the existing literature on reported utility, we control for prices, hence taking into account the critical role of *value for money* when judging services.

The remainder of the paper is structured as follows: the relevant streams of literature, focusing on the role played by expectations and surprise on individual utility, are recalled in Section 2, while the theoretical framework is presented in Section 3. We then describe the data and the empirical strategy (Section 4). Finally, the results and the concluding remarks are presented in Sections 5 and 6, respectively.

2. Literature review

Our research contributes to the growing body of literature that explores the significance of expectations in economics, particularly focusing on the impact of discrepancies between anticipated and actual outcomes on individual utility, a phenomenon referred to as a *surprise*.

2.1. Surprises in economics

Surprises are directly related to expectations and uncertainty about future outcomes. By definition, surprises emerge as a conflict with pre-existing beliefs (Baccan et al., 2015). From a psychological perspective, according to Schützwohl (1998), surprises result from schema-discrepancy, *i.e.*, an inconsistency between activated schemas and newly acquired information. When such discrepancy exceeds a certain threshold, people experiment with the feeling of surprise, which undergoes a certain degree of subjectivity because people might display different thresholds. In economics and financial studies, the surprise can be defined as the mismatch between expected and actual results, for instance, when the rate of return of investments is considered.

Economic decision-making is inherently uncertain and relies on expectations, which in turn impact the utility derived from consumption. While adherence to prior expectations is important for utility, a mismatch does not necessarily have negative connotations. Research has shown that the emotional response to surprise can be either positive or negative (Janakiraman et al., 2006). The 'sign' of the surprise depends on each agent's consequence of the unexpected event or outcome. As discussed in Derbaix and Vanhamme (2003), the level of surprise associated with an event can amplify the reaction to it. This means that unexpected events will generally elicit stronger responses compared to equally significant but expected ones.

Most of the research dealing with the effect of surprises on people's utility and enjoyment is focused on entertainment markets, such as sports games (Bryant et al., 1994; Ely et al., 2015; Peterson and Raney, 2008) or online media products (Simonov et al., 2022). Extant literature underlines the importance of both suspense (*i.e.*, excitement or anxiety about what will happen in the future) and surprise dynamics in the context of entertainment. However, it is hard to empirically assess their role because beliefs and enjoyment are often not directly observable (Ely et al., 2015). In this regard, Bizzozero et al. (2016) study preferences using the size of the tennis audience as a proxy for enjoyment with a match. As a matter of fact, bored viewers can easily (at almost no cost) decide to switch the channel or turn the TV off. While in these specific markets' enjoyment is expressed in terms of audience, when dealing with the consumption of other services, it can be assessed via consumers' explicit evaluations in the form of online reviews (Fradkin et al., 2018). In an experimental setting, Derbaix and Vanhamme (2003) studied the effect of elicited surprises on the willingness to share information (word-of-mouth).

2.2. Exogenous shocks and individual utility

While most of the economic attention is directed toward the impact of an exogenous shock on the macroeconomy, the effect of shocks on agents' utility is a relatively unexplored topic in the microeconomics literature; this fact most probably results from standard economic theory, which uses probability theory to treat expectations and uncertainty objectively. At the microeconomic level, Oswald and Powdthavee (2008) build upon the hedonic adaptation theory (Frederick and Loewenstein, 1999) to assess the effect of onset disability on people's well-being. This term refers to the way individuals react and adapt to upcoming favourable or adverse events, which, in the context being discussed, can be defined as exogenous shocks.

According to psychology literature, shocks' effect on people's utility is not permanent, *i.e.*, reactions fade over time because people tend to adapt to new conditions. This is what Wilson and Gilbert (2008) defines as affective adaptation. Rayo and Becker (2007) equate hedonic utility to happiness and look at its volatility over time. In this regard, they consider that happiness is not constant and depends on upcoming changes, prior expectations about the future level of happiness, and peers' happiness. In a similar vein, Kettlewell et al. (2020) analyse the well-being time path of women affected by an exogenous shock, namely the spouse's death. Riis et al. (2005) analyse the reported level of happiness of healthy people and people with serious illnesses, finding no evidence of a significant difference. All these studies align with the theory of hedonic adaptation, showing that, after an external shock (which might be of any nature), utility tends to converge to the pre-event level. Moreover, the empirical evidence on hedonic adaptation also holds for other types of shocks, more related to individuals' disposable income. In this regard, Easterlin (1995) finds evidence that an income increase does not lead to long-run higher well-being. This is explained in terms of the hedonic treadmill: people's happiness does not increase with money since people tend to quickly adapt to the new income level.

A classic example of a purely exogenous shock is the one stemming from weather-related events. Weather affects everyday life, influencing and limiting the range of activities that can be performed. Weather conditions might influence people's choice between work and leisure (Connolly, 2008) and their overall productivity (Lee et al., 2014). When choosing between leisure and work, bad weather could decrease the opportunity cost of working (reducing distractions resulting from good weather), hence bringing more work hours or, on the contrary, increase it when workers have a daily income target that can be reached earlier when the weather is bad, as in the case described by Camerer et al. (1997) of New York cab drivers who work more intensively in rainy days. In a similar fashion, Huysmans (2002) looks at the weather as a determinant of human sleeping, sports participation, leisure, and other recreational activities. The weather might also moderate the participants' responses to financial events (Dehaan et al., 2017). Finally, when buying valuable or durable goods, consumers must forecast how much utility they will derive from future consumption, which might take place in different states of the world. In such conditions, consumers are exposed to a variety of psychological biases such as projection bias and salience effects: Busse et al. (2015) find that the choice to purchase a convertible or a four-wheel-drive is highly dependent on the weather at the time of purchase in a way that is inconsistent with classical utility theory. Similarly, Conlin et al. (2007) find evidence of projection bias over the weather: specifically, people's decisions in buying cold-weather items are overinfluenced if the current weather is cold.

Scaling down to the context of the current study, the weather undoubtedly has remarkable effects on travel choices and enjoyment of related leisure activities. This relationship exists for practical reasons; for instance, people tend not to go to the beach when raining or visit cities in the middle of a heatwave. Moreover, from a more psychological perspective, a stream of literature links the weather with mood, thinking, and judgments (Klimstra et al., 2011). In a recent paper, Brandes and Dover (2022) look at the effect of unpleasant weather (hence assuming preferences for sunny weather) on evaluating accommodation services. More specifically, they consider the weather in the origin country (which does not coincide with the destination country) and find that rain is associated with a higher review provision and lower scores. This result is most likely due to the lower opportunity cost of time when raining and the bad mood associated with bad weather. Similarly, Bujisic et al. (2019) show that weather factors such as rain, temperature, and barometric pressure drive consumers' complaint behaviour in restaurants.

While most of the literature has focused on the effect of weather *per se* on the enjoyment of leisure activities (Gössling et al., 2016; Jeuring, 2017; Jeuring and Peters, 2013), less is known about the effect of weather surprises on agents' utility. In other words, existing studies do not consider how the mismatch between agents' expectations about the weather and the real conditions when travelling affects their mood and, hence, their evaluation of the experience. This is precisely where our work is positioned. On these premises, it is reasonable to think that the effect of weather status on agents' utility also depends on the degree of the unexpectedness of such status (Derbaix and Vanhamme, 2003), *i.e.*, on the mismatch between existing information at the time of expectations building (for instance, derived from weather forecasts) and the realized weather. As noted by Figini et al. (2022), weather forecasts are public information that significantly impacts travel decisions and firms' pricing strategies, despite being fully exogenous. In the case of weather forecasts, this information flow incurs negligible costs, thereby not being subject to potential rational inattention (Sims, 2006).

3. Theoretical framework

We assume that visitors, the economic agents under investigation, hold a belief about the weather that will occur during their consumption experience, the holiday. This expectation affects the type of activities they plan to carry out and, in the first place, the very decision to travel to a specific destination at a specific time of the year. Any deviation from the expected weather conditions can disrupt their anticipated activities, leading to a non-neutral effect on utility and enjoyment. Although weather is unrelated to the intrinsic service being consumed and reviewed, we investigate if such non-fundamental shocks can be reflected in online rating scores, thereby introducing an informational distortion into the metric used to assess the quality of the service.

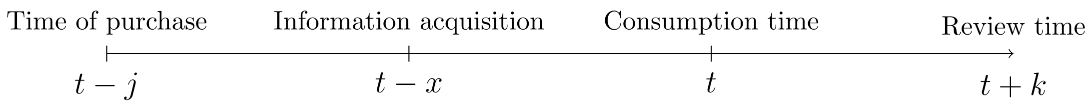

Fig. 1. The consumption timeline.

In our setting, a key assumption is the lack of *rational inattention* (Sims, 2006). Since forecasts are publicly available information that is accessible to all agents at little to no cost, subjects can easily adjust their expectations based on weather forecasts published in the days before travelling. Agents plan the activities to be carried out during the consumption experience based on their beliefs about future weather, which stem from historical climate conditions, weather trends and, if available, forecasts at the time of booking. Weather forecasts are built on meteorological models and are used to predict the realization of the weather on a specific date. Forecasts share some degree of inaccuracy, and they become more accurate as the date to be forecast approaches. The *surprise* results if the realized weather does not match the weather forecast. It is important to note that also the weather *per se* might affect the experienced utility, as agents purchase a service to be consumed in a specific period and are aware of the ex-ante probability of having different types of weather while at the destination. In other words, when booking for specific dates, agents have already factored in the expected utility derived from the possible states of the world, weighted by their probability based on historical data (and common sense).

In its simplest form, we can model the weather as a qualitative nominal variable: it can either be Rainy (R), Cloudy (C), or Sunny (S). Given this theoretical context, four specific points in time are relevant for the purpose of our study:

1. The time of purchasing, which occurs in advance with respect to the consumption time;
2. The time preceding consumption when the agent acquires the latest available information (the most accurate weather forecast)[2];
3. The time of consumption, which corresponds to the delivery of the purchased service (and the realized weather)[3];
4. The time of review, when the online rating is provided.

To facilitate understanding of the time gap between booking and review time, the timeline is graphically represented in Fig. 1. If consumption occurs at time t, the review is written at time $t + k$ (where $k > 0$). Going backwards, the purchasing occurs at time t-j (where $j > 0$). Between t-j and t, the agent receives updates about the weather forecast for time t. As the consumption event approaches, these signals become increasingly more accurate. Therefore, we assume that on the eve of consumption (t-x) (where $x < j$), the agent acquires a trustworthy signal that serves as the basis for their expectations.[4]

As regards the weather signal, let us denote with Ω the finite space of states ω. $\omega \in \Omega = S, C, R$, where S denotes a sunny day, C denotes a cloudy day, and R represents a rainy day. At the time of purchasing the service, the agent forms a belief μ, with μ^{ω} being the probability of the state ω. μ_0 is the prior, *i.e.*, the ex-ante likelihood of ω to occur before considering any new (posterior) information. We expect that for values of j big enough (early bookings), the agent does not avail of reliable forecasts so that the expectations are made on past trends (for example, the historical weather trend for the same period of the year and common sense). On the consumption eve, the agent receives a signal w (weather forecast), which produces expectations about the activities that can be performed.[5]

Given the received signal w, and its adherence to the realized ω, we have the scenario represented in Table 2. The surprise appears if w is incorrect (or, alternatively, $w \neq \omega$). Intuitively, the intermediate state C, representing a cloudy day, might arguably be associated with a *weak surprise* while the sudden change from sunny to rainy, or vice-versa, might constitute a *strong surprise*. Hence, Table 2 shows two cases of *strong surprise*, four cases of *weak surprise*, and three cases of *no surprise*. *Ceteris paribus*, we expect that weather surprises affect agents' experienced utility and can be captured in online rating scores. Additionally, according to the good-news bad-news theory (Eil and Rao, 2011), the effects of S→R and R→S may not be symmetrical, with the magnitude of these effects potentially differing for positive (good news) and negative (bad news) shocks.

Before discussing the empirical setting, one crucial point should be mentioned. The concept of utility used in the current work is not Decision Utility, used in standard economic literature to explain agents' choice, but Experienced Utility (Kahneman and Thaler, 1991). As discussed by these authors, Experienced Utility is measurable and has a neutral point on the boundary between desirable and undesirable. This is indeed more fitting in our study, which recalls utility as the evaluation of an episode. More specifically,

[2] Note that points 1 and point 2 could collapse into one timeframe in the case of late bookings.

[3] Furthermore, the consumption period, t, varies in length and can also be represented as $t(t_1, t_2)$, where t_1 is the initial day of the consumption (the check-in day) and t_2 marks the end of the consumption (the check-out day).

[4] It is common for individuals to consult the weather forecast prior to embarking on a trip, as it enables them to be better equipped for any weather contingencies that may arise, such as rain or extreme temperatures. In a straightforward manner, this information is utilized to make informed decisions on what attire and personal belongings to pack for the trip.

[5] For simplicity, we hold the expected quality of the accommodation service constant. Agents choose the service provider based on a set of characteristics and reviews from previous guests (Acemoglu et al., 2022). However, one might expect that online reviews (which are other types of signals) posted between purchase and consumption could generate a path of evolving expectations about the hotel quality and hence affect future enjoyment. In line with existing literature, we expect that for verified platforms, such as Booking.com, ratings tend to be quite stable over time (Figini et al., 2020). Moreover, consumers tend to read reviews and gather information only before booking accommodation to ensure they make an informed decision.

Table 2
The surprise matrix.

		Realised (ω)		
		S	C	R
Forecast (w)	S	No surprise	Weak surprise	Strong surprise
	C	Weak surprise	No surprise	Weak surprise
	R	Strong surprise	Weak surprise	No surprise

among the different types of experienced utility, we link to the specific concept of remembered utility, which consists of a '*measure of hedonic and affective experience, inferred from a subject's retrospective reports of the total pleasure or displeasure associated with past outcomes*' (Kahneman and Thaler, 2006).

Lastly, we want to emphasize that the weather can be regarded as instrumental information in determining the decision to purchase accommodation services, while sudden weather changes are non-instrumental information in this respect. Unforeseen weather changes may have an impact on overall satisfaction with the stay, but they cannot directly influence the decision to book the accommodation, which is made prior to arrival. In fact, it is only during consumption that the surprise becomes apparent and has the potential to impact the experience.

4. Data and methods

4.1. Data

Since we aim to investigate whether and to what extent the subject's reaction to surprises (sudden changes in weather conditions) is reflected in their post-experience evaluation, we use the individual rating score as a proxy for experienced utility. We retrieve individual ratings and other characteristics of the stay from Booking.com, a popular hotel reservation platform.[6] Weather forecasts are collected from various sources, including private providers such as AccuWeather, as well as local public providers such as Arpa Lombardia and Arpa Veneto, the regional environmental protection agencies that are required by Italian law to produce and publish weather forecasts for the next 72 h. Our study focuses on two Italian cities, Venice and Milan, covering the period between September 2019 and February 2020. Italy is one of the most visited destinations in the world, offering a diverse range of activities. By selecting these two cities, we aim to explore contexts mainly characterized by culture and leisure (Venice) and culture and business activities (Milan).

The study period allows for a fair comparison of experienced utilities since all scores have been generated under the same review system. In fact, Booking.com changed its rating algorithm starting in September 2019. In the new system, reviewers are asked to provide a numerical rating (from 1 to 10) of the overall experience. Unlike the previous system, which requested reviewers to evaluate only six specific hotel items through smileys (while the overall score was generated as a simple average), the current overall score is explicitly rated by the customer and could arguably better reflect the influence of exogenous factors that are beyond the control of the hotel. Additionally, we stopped the data collection in February 2020 to avoid any confounding effects generated by the COVID-19 pandemic and subsequent lockdown measures.

4.1.1. Data collection and descriptive statistics

Accommodation. To collect the accommodation data, we used an ad-hoc web crawler that emulated online users' browsing behaviour. The crawler recorded individual ratings and prices for any room offered on the platform by the population of 3- to 5-star hotels in the two cities under scrutiny. During the six-month period, we collected a total of 341,039 observations, each consisting of an individual numerical score of the accommodation service for a specific structure and travel date.

For each observation, we have data on (i) the accommodation characteristics, (ii) the date when the review was posted, (iii) the month of the stay, (iv) the consumer's profile, including name, country of origin, travel party, and length of stay at the hotel, (v) the pre-existing score, which is the average score computed on the stock of previous reviews up to the week before the stay, and (vi) the price. This last variable represents a precious piece of information, considering that empirical studies investigating online reviews tend to overlook prices, hence not taking into account value-for-money judgments and the role of price in shaping consumers' expectations. However, it is important to highlight that we do not have access to information regarding the exact price associated with each stay (and review). This is because the price of a given hotel room can vary depending on various factors stemming from dynamic pricing strategies, including the time between the booking date and the stay date, and additional fares, like the free cancellation/non-refundable options and the breakfast inclusion/exclusion. To account for this variability, we built a

[6] The investigation of how textual reviews mention external factors (e.g. weather) and related surprises might also be of interest, although not considered here. However, given our theoretical framework for which the surprise works as an unconscious factor, we expect very little mentioning of surprises related to external factors in textual reviews.

Table 3
Summary statistics of Booking.com variables.

Label	Description	Mean/Percentage	St. Dev.	Min	Max
`Score`	Individual score (outcome variable)	8.430	1.643	1	10
`LengthOfStay`	Length of the stay (in days)	1.817	0.771	1	3
`Couple`	=1 if travel party = couple	0.456			
`Family`	=1 if travel party = family	0.260			
`Group`	=1 if travel party = group	0.087			
`SoloTraveler`	=1 if travel party = solo traveller	0.125			
`Domestic`	=1 if Italian	0.260			
`Anonymous`	=1 if anonymous reviewer	0.135			
`LnPrice`	Average price (in ln)	5.361	0.641	2.944	8.422
`PriorRatings`	Average of pre-existing score	8.482	0.624	4.450	10

measure of price for each observation by computing the simple weekly average of the different prices posted by the hotel for any room type and for bookings with different advances.[7]

Table 3 presents the descriptive statistics for the relevant variables. The individual score, which serves as the dependent variable in our study, proxies the experienced utility of the stay and ranges from 1 to 10. The mean score is 8.43, and the variable is highly left-skewed, with approximately 26% of observations receiving a score of 10/10. This reflects the classic J-shaped distribution of online reviews (Hu et al., 2017; Li and Hitt, 2008). 67% of reviews are for providers located in Venice (125 structures), while the remaining 33% is for Milan (47 structures). This imbalance in the geographic distribution of observations is attributable to the higher number of stays in Venice compared to Milan (ISTAT, 2020).[8] Moreover, the type of consumers staying overnight in the two cities is quite different. As a matter of fact, Venice hosts a higher share of leisure visitors, while Milan is mainly known for business travellers.

On average, each provider received approximately 3240 reviews in the period under investigation, with a high degree of heterogeneity across providers (min = 20; max = 8262; St. dev. = 1926.43). 26% of reviewers are Italians; 45% travelled as a couple, 26% as a family, 12% are solo travellers, while groups represent the remaining 13%. 14% of scores are from anonymous customers. However, as per the platform policy, only real guests can rate the service, which ensures high reliability in this metric. The average number of nights spent by consumers in the accommodation structure is 1.8.[9] The average price per night is €204, displaying a high heterogeneity across dates and providers (min = €50.40; max = €4491.76).

Weather. Data on weather forecasts were collected from the AccuWeather website, owned by an American media company providing commercial weather forecasting services worldwide. For a robustness check, we have also complemented data with the weather forecasts and the official data on weather conditions recorded by the Regional Environmental Protection Agency of the two involved regions (Arpa Veneto for Venice and Arpa Lombardia for Milan).

As reported in Table 4, there was unexpected rainy weather (*i.e.*, the forecast announced sunny weather while it actually rained) in around 6% of the days under investigation. If we also consider the cases in which the forecast announced cloudy weather, but it actually rained (what we label as a `NegativeSurprise`), the percentage increased to 16%. Analogously, 0.3% of days had unexpected sunny weather (*i.e.*, the forecast announced rainy weather while it was actually sunny), increasing to 3.1% if we also consider forecasts of cloudy weather which turned out to be sunny (what we label as `PositiveSurprise`). This set of dummy variables has been created starting from a discrete variable generated using data from AccuWeather. This variable describes the daily weather according to the phrase and icon that the provider presents on its website and app. It has been consequently categorized into three levels: sunny, cloudy, and rainy. The use of weather icons, as opposed to more detailed information about humidity, air pressure, solar radiation, etc., makes weather forecasts easier for final users to interpret and, hence, further reduces the mental effort required to decode the information.

To perform robustness checks, we also created a dummy variable, `sunniericon`, which captures situations where the actual weather was better than the forecast weather (e.g. if the forecast was rainy and the actual weather was cloudy). Additionally, we generated two continuous variables representing the differences between the actual and forecast sunny hours `DiffHoursSun` and the actual and forecast probability of rain (`DiffRainProb`). For brevity, we have included the descriptive statistics and further details on these variables in the Appendix (Table A1-A3).

[7] To clarify, we computed the price by tracking the price trajectory for a specific room type over time. We did this by collecting the different prices posted by the hotel for that specific room type at different time points, starting from 15 days before the travel date and up to the actual travel date. We then computed the mean of these prices at each time point and averaged these mean values over the course of one week. This resulted in a single price value for each observation, which represents the average price associated with a stay in the hotel over a one-week period. This approach helped to average out differences that could arise from varying booking windows.

[8] Official Statistics on Tourism Flow, Istat. Website: https://www.istat.it/it/files//2020/12/C19.pdf. Accessed on December, 13th 2022.

[9] Data have been restricted only to short stays (1–3 days). The reasons behind this choice are discussed later in this Section.

Table 4
Summary statistics weather data.

Label	Description	(%)
UnexpectedRain	Forecast: sunny → rainy	0.063
NegativeSurprise	Forecast: sunny or cloudy → rainy	0.164
UnexpectedSun	Forecast: rainy → sunny	0.003
PositiveSurprise	Forecast: rainy or cloudy → sunny	0.031
Sun	Sunny weather	0.353
Rain	Rainy weather	0.266
Cloudy	Cloudy weather	0.381

4.2. Sample construction

Our dataset integrates data from Booking.com (ratings and accommodation characteristics) with those on the weather (forecast and actual weather). It is worth noting that while the exact date of the review is available on the platform, only the length of stay, the month and the year of the rated stay are provided, thereby creating a potential issue when matching the review with weather data. To minimize the error, we have thereby limited our analysis only to short stays (between one and three days) and only to those reviews where the month of the review and the month of the stay coincide.[10]

Unlike Brandes and Dover (2022), who investigated undisclosed platforms different from Booking.com and found that reviews are written on average 3.5 days after the stay, we adhere to Booking.com's policy for which, during the period under investigation, customers could post their reviews when answering a specific e-mail sent by Booking.com two days after the check-out. We assume that customers rated soon after receiving the email or did not rate at all. Next, we assume that consumers form their expectations when accurate weather forecasts are available. Since the accuracy of weather forecasts is 95% in the last 72 h[11] and given the restricted sample (we only considered short stays between one and three days), we assumed that weather checks are made two days before departure. Recalling the notation of Fig. 1, given the stay $t = [t_1; t_2]$, $k = 2$, and $x = 2$.

On these premises and pondering that the length of stay ranges between 1 and 3 days, we first consider the average length of stay of 2 days, with the weather forecast being checked two days prior to the start of the stay and reviews written 2 days after the end of the stay. For instance, if the review date is October 10th, we assume that the customer ended their stay on October 8th. Thus, we match each review with the realized and forecast weather for the check-in date (*i.e.*, October 6th), which was posted on October 4th. We match each stay date with the corresponding weather conditions of the city where the accommodation is located (Venice or Milan), and we then build the surprise variables by comparing, for the check-in date, the weather conditions and the weather forecast produced two days before. It is important to highlight that, to avoid further complications in the matching between reviews and stays, the surprise is measured only on a single day, the initial day of the short stay. A similar procedure is repeated by matching the review date and the weather forecast, taking into account the actual length of stay attached to each review. The corresponding estimates, produced for robustness check, can be found in the Appendix (Table A4).

4.3. Empirical strategy

The effect of surprises on experienced utility is estimated through the following utility model:

$$U_{ijh(t+k)} = \beta' \boldsymbol{X_i} + \mu' \boldsymbol{Z_{ht}} + \gamma' \boldsymbol{Weather_{jt}} + \tau' \boldsymbol{Surprise_{j(t;t-x)}} + \alpha_h + \alpha_m + \epsilon_{ijht} \tag{1}$$

where $U_{ijh(t;t+k)}$ is the utility reported at time $t + k$ by consumer i who consumed the accommodation service of provider h, located in the city j, at time t. $\boldsymbol{X_i}$ is a vector of consumer characteristics (nationality, anonymity, length of stay, and travel party). $\boldsymbol{Z_{ht}}$ is a vector of time-varying provider characteristics (price, in logarithm, and the stock of previous reviews). $\boldsymbol{Weather_{jt}}$ and $\boldsymbol{Surprise_{j(t;t-x)}}$ are two vectors including a series of dummies associated to, respectively, the realized weather at time t in city j and the weather forecast for time t produced at time t-x. γ' and τ' are the main parameters to be estimated, capturing the effect on the outcome of the weather and of the surprise. α_h are provider-fixed effects, which address issues of time-unvarying unobserved heterogeneity across different providers. α_m are monthly fixed effects, controlling for seasonal variations and hence for prior beliefs on the weather, and ϵ_{ihjt} is the stochastic error term.

Based on hedonic adaptation theory, we expect that the effect of exogenous shocks weakens for longer stays since the effect (recall it is measured on the first day) gets diluted or spread over time. To account for this, we provide an alternative specification that includes interaction terms between surprise and stay duration (expressed in days; remember that we only consider short stays between one and three days). More formally:

$$\begin{aligned} U_{ijh(t+k)} &= \beta' \boldsymbol{X_i} + \mu' \boldsymbol{Z_{ht}} + \gamma' \boldsymbol{Weather_{jt}} + \tau' \boldsymbol{Surprise_{j(t;t-x)}} \\ &\quad + \tau' \boldsymbol{Surprise_{j(t;t-x)}} \times LengthOfStay_i + \alpha_h + \alpha_m + \epsilon_{ijht} \end{aligned} \tag{2}$$

[10] We acknowledge that for reviews written in the first two days of the month, the reported stay month may differ from the month of the review. For example, if a guest checks out from a hotel on October $31st$, the review will likely be written in the early days of November. We have taken great care to handle such cases appropriately in our analysis.

[11] According to the Aeronautica Militare webpage: https://www.meteoam.it/it/attendibilita-previsioni. Accessed on January, 27th 2023.

The models specified in Eqs. (1) and (2) are estimated via Ordinary Least Squares with a robust variance estimator.[12]

5. Results

5.1. Main findings

Table 5 shows the results of the Model of Eqs. (1) and (2). Our main finding is that surprises arising from the mismatch between expected and realized weather systematically affect the rating score, thereby modifying customers' experienced utility. More specifically, `unexpectedrain` is associated with a significant decrease in the overall utility (−0.150***, Column 1) when the rain comes as a surprise during the stay, *ceteris paribus*. Although the drop's magnitude might appear negligible, we recall that the dependent variable ranges on a 1 to 10 scale, with a relatively low variation around the mean (st.dev. = 1.643). Surprises with the opposite sign (`unexpectedsun`, that is, sunshine when forecasts announced rainy weather) translate into higher experienced utility (0.337***, Column 1). Notice that Model 1 controls for realized weather conditions (`Sunny`, `Rainy`, with `Cloudy` excluded as the reference category) to avoid the surprise capturing the effect of realized weather conditions. Noticeably, the coefficients of both `Rainy` and `Sunny` are always significant, with a negative sign (contrary to common sense in this latter case, we will return on this result below).

A potential source of heterogeneity in the effect's magnitude is the consumption span: it is likely that the experience's length might dilute the surprise effect. Consistently with the model of Eq. (2), we expect that the surprise effect captured by reviews diminishes with the length of stay. The interaction between the surprise variables with the number of days spent in the city (`UnexpectedRain#LengthOfStay`; `UnexpectedSun#LengthtOfStay`) supports the theoretical construct: the effect weakens for longer stays (0.125***; −0.125**, respectively, Column 2). In accordance with the hedonic adaptation theory (Frederick and Loewenstein, 1999), longer stays arguably provide more time for adaptation to unexpected circumstances, resulting in a more diluted surprise effect that spreads over a longer period.

Our results also show the asymmetry in reactions to different types of surprises, which is consistent with studies on asymmetric responses to financial uncertainty. Extant literature, theoretically grounded in the good news/bad news theory, consistently finds that negative information has a much greater impact on individuals' attitudes than positive information does (Eil and Rao, 2011; Nguyen and Claus, 2013; Soroka, 2006). Whether our findings support the existence of a negativity bias (Schwager and Rothermund, 2013) is open to discussion. On the one hand, estimates suggest that the increase in reported utility when having unexpected good weather is higher than the drop caused by unexpected unfavourable weather. On the other hand, the negativity bias is typically associated with information disclosure which is, in our case, the weather forecast. Hence, the bad news associated with the rain forecast changes the subjects' reference point (decreasing it), while the good news of sun forecast does not change the reference point because is consistent with expectations at the time of booking. Hence, the higher absolute values of the coefficients associated with positive surprises in all the specifications of Table 5 can be interpreted as the over reaction that discounts the previous negative change in the reference point, consistently with the negative bias discourse.

In Column 3, we test the same model specification, but this time, we only consider days with expected or realized sunny or rainy weather. Therefore, we exclude cloudy days from both the forecast and actual observations (`RAINY` becomes the reference category). This is done to check if the results are driven by the "grey area" represented by cloudy weather, where arguably the single consumers' reaction might be more heterogeneous. However, the findings of Column 3 are similar to the full model.

In Column 4, we examine a different specification of the model, where we test the impact of receiving `PositiveSurprise` (forecast: cloudy or rainy; realized: sunny) or `NegativeSurprise` (forecast: cloudy or sunny; realized: rainy) compared to the reference category, which is correctly predicted weather. The results confirm that, compared to correctly predicted weather, experiencing a positive surprise boosts utility significantly (0.0794***), while experiencing a negative surprise translates into a lower reported level of satisfaction (−0.0535***). Again, the magnitude of the positive surprise is stronger than that of the negative counterpart, this result being consistent with the literature on the negativity bias, as the positive surprise is the overreaction to the bad news received with the weather forecasts, which decreased the expectations' reference point.

In Column 5, we test a slightly different specification of Model (1), which restricts the analysis to rainy days. In other words, conditional of being rainy, we look at the effect of a strong surprise (`S2R` = sunny to rainy) and a weak (`C2R` = cloudy to rainy) surprise, with no surprise (`R2R` = rainy to rainy) as the reference category. Similarly, conditional on being sunny, Column 6 studies the effect of strong (`R2S` = rainy to sunny) and weak (`C2S` = cloudy to sunny) surprises, with a correctly predicted sunny weather `S2S` (= sunny to sunny) as the reference category. Results show that, compared to a rainy day with no surprise (*i.e.*, the rain was correctly forecast), there is a significant and negative effect on the rating score of the unexpected rain when the forecast was of a sunny day (−0.129***, Column 5). Interestingly, the effect is much weaker (statistically nil) when the unexpected rainy weather follows a forecast of cloudy weather only (`cloudy to rainy`, Column 5). In the same fashion, there is a strong increase in the reported rating scores when agents experience sunny weather during their stay after a wrong forecast of rain (0.522***, Column 6). Again, such an effect is weaker when the forecast is of cloudy weather (0.192***, Column 6), as expected.

In Column 7, we run the fully specified model, which includes the 8 dummy variables, with the reference category being the correctly predicted cloudy weather (`C2C`). Results again confirm the effect of strong surprises (i.e., `R2S` and `S2R`) while cloudy weather has less clear and precise directions, coherently with the theoretical framework and with the intuition that people might

[12] In the main estimations, we use Huber-White heteroskedasticity-robust standard errors.

Table 5
Pooled OLS estimation of models in Eq. (1) and in Eq. (2).

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
LengthofStay	−0.0248***	−0.0316***	−0.0210***	−0.0243***	−0.0235***	−0.00638	−0.0247***	−0.0258***
	[−6.76]	[−8.36]	[−4.22]	[−6.63]	[−3.09]	[−1.05]	[−6.72]	[−5.03]
Domestic	−0.0426***	−0.0433***	−0.0627***	−0.0427***	−0.0327**	−0.0995***	−0.0422***	−0.0440***
	[−6.46]	[−6.57]	[−7.24]	[−6.46]	[−2.33]	[−9.11]	[−6.40]	[−4.81]
Couple	0.0168**	0.0175**	0.0357***	0.0192**	−0.0174	0.0735***	0.0200**	0.0387***
	[2.03]	[2.12]	[3.32]	[2.33]	[−1.04]	[5.21]	[2.42]	[3.33]
Family	0.0469***	0.0462***	0.0589***	0.0480***	0.0973***	0.0449***	0.0491***	0.0586***
	[5.23]	[5.15]	[5.05]	[5.35]	[5.23]	[2.99]	[5.47]	[4.66]
Single	−0.0005	−0.0006	−0.0191	0.00222	−0.0996***	0.0649***	0.00126	0.00866
	[−0.05]	[−0.06]	[−1.39]	[0.21]	[−4.58]	[3.69]	[0.12]	[0.59]
Anonymous	−0.203***	−0.202***	−0.217***	−0.201***	−0.0616***	−0.313***	−0.201***	−0.258***
	[−25.68]	[−25.61]	[−21.25]	[−25.44]	[−3.98]	[−22.94]	[−25.37]	[−23.04]
LnPrice	−0.267***	−0.269***	−0.210***	−0.273***	−0.389***	−0.107***	−0.263***	−0.171***
	[−16.22]	[−16.36]	[−9.40]	[−16.58]	[−10.03]	[−3.92]	[−15.92]	[−7.41]
PriorRatings	−2.031***	−2.026***	−1.649***	−2.041***	−1.612***	−2.267***	−2.019***	−1.859***
	[−27.23]	[−27.14]	[−17.90]	[−27.45]	[−11.46]	[−18.22]	[−27.13]	[−16.64]
Sunny	−0.0624***	−0.0626***	−0.0395***	−0.101***				
	[−9.87]	[−9.89]	[−3.94]	[−11.77]				
Rainy	−0.0192**	−0.0185**		−0.0556***				
	[−2.33]	[−2.26]		[−7.37]				
UnexpectedRain	−0.150***	−0.379***	−0.368***					
	[−10.30]	[−11.79]	[−11.18]					
UnexpectedRain#LengthOfStay		0.125***	0.119***					
		[8.48]	[7.88]					
Unexpectedsun	0.337***	0.540***	0.600***					
	[10.00]	[5.45]	[6.03]					
Unexpectedsun#LengthOfStay		−0.125**	−0.157***					
		[−2.24]	[−2.80]					
PositiveSurprise				0.0794***				
				[6.55]				
NegativeSurprise				−0.0535***				
				[−6.89]				
R2R							−0.0324**	
							[−2.42]	
R2C							−0.0812***	
							[−3.85]	
R2S						0.522***	0.239***	
						[11.18]	[6.96]	
C2R					0.0102		−0.0882***	
					[0.72]		[−7.10]	
C2S						0.192***	0.0209	
						[11.33]	[1.17]	
S2R					−0.129***		−0.221***	
					[−7.07]		[−14.07]	
S2C							−0.0605***	
							[−5.90]	
S2S							−0.126***	
							[−12.22]	
Diff.Icon								0.0141***
								[4.94]
Constant	27.16***	27.14***	23.63***	27.32***	24.06***	28.28***	27.10***	25.21***
	[42.71]	[42.65]	[29.93]	[43.08]	[19.71]	[26.47]	[42.68]	[26.43]
N	320 685	320 685	199 269	320 685	79 957	119 312	320 685	165 377
MONTH FE	YES	YES	YES	YES	YES	YES	YES	YES
PROVIDER FE	YES	YES	YES	YES	YES	YES	YES	YES
adj. R2	0.147	0.147	0.146	0.147	0.173	0.159	0.147	0.152

Note: The reference category for Weather is Cloudy in (1), (2) and (4); Rainy in (3). The reference category for Surprise is No surprise in (4); R2R in (5); S2S in (6); C2C in (7). Estimates in columns (1), (2), (4), and (7) pertain to the whole sample. Column (3) excludes cases with forecast or actual cloudy weather. Columns (5) and (6) focus on rainy days and sunny days, respectively. Column (8) includes only days without meteorological surprises.

have a less generalized taste for cloudy weather. Interestingly, while it could be reasonable that rainy, even if well predicted (R2R), is less desirable than cloudy weather, S2S displays (like in Columns 1 and 3) a negative and significant coefficient, which will be robust in any alternative specification (Appendix Section). To delve deeper into this matter, we considered how surprises are categorized. In the current work, surprises are represented by dummy variables using three icons to summarize different states of the world. However, weather conditions are more nuanced than this simple framework suggests, with various degrees of sunshine (e.g., moderately sunny, sunny, very sunny) and the same for rainy and cloudy conditions. This simplified categorization treats all these variations as one category. For instance, if the weather forecast is mild rain and the actual weather turns out to be a thunderstorm, in our setting, it is labelled as a NoSurprise scenario despite containing a modest negative surprise. Therefore, in Column 8, we exclusively consider scenarios with purportedly no surprises (R2R; C2C; S2S) and introduce a variable named DiffIcon. This variable captures the discrete transition from one state of the world to another, ranked from the worst (worst rainy weather) to the best (very sunny weather), encompassing both positive and negative values.[13] Results reveal that an increase in the DiffIcon value, indicating a transition from one status to another, more positive, within the same icon (e.g., from partially sunny to very sunny), correlates with higher reported utilities.

5.2. Further results

Although the main goal of the work is to isolate the effect of the exogenous surprise from other confounding elements that also affect the rating score, this sub-section offers a quick overview of the estimated coefficients of the other covariates. Estimated coefficients in Table 5, despite some marginal differences in their magnitude across specifications, remain consistent. As expected, price (lnprice) negatively impacts scores: for any given hotel, when average prices are higher, ratings are lower, capturing the impact of *value for money*. Interestingly, anonymous reviews are associated with lower scores, a result that aligns with deindividuation theories. As Deng et al. (2021) explain, anonymity enables reviewers to give worse ratings because of a lower self-awareness and social presence. Domestic (*i.e.*, Italian guests) tend to provide lower scores than foreigners, perhaps due to a better familiarity with the reference quality that can be expected from the service (Cordell, 1997). Moreover, couples and families are associated with higher scores compared to people travelling in groups.

Longer stays (LengthOfStay) are associated with lower scores: it is reasonable to think that spending more time in the accommodation, people might be more analytical about the evaluation of the services (Kim and Han, 2022) and that longer stays decrease the comfort sensation with hotel facilities. Another interesting result concerns the effect of PriorRatings. As discussed in the theoretical framework, the nature of the analysed product as an experience good (*i.e.*, a good whose quality can be ascertained only upon consumption) makes other agents' opinions an important source of information when choosing a service provider. Our results suggest that, everything else being equal, a higher average rating translates into lower individual scores, perhaps due to the internalization of higher expectations about quality.

5.3. Robustness checks

We conducted a series of robustness checks to assess the sensitivity of our findings. Initially, in Columns 1–3 of Table A4 (Appendix Section), we re-estimate the specifications reported respectively in Columns 2, 4, and 7 of Table 5 employing an alternative approach to compute the surprise date. In the main model, of Table 5 we subtract 2 days (the average length of stay) from the checkout date to obtain the surprise date. In this alternative specification of Table A4, we subtract the reported length of stay, which can vary from one to three days in our sample. In Columns 4–6 of the same table (Table A4, Appendix), we re-estimate again the specifications reported, respectively, in Columns 2, 4, and 7 of Table 5, this time with bootstrapped standard errors to consider the estimators' variance. Finally, in (Columns 7–9, Table A4, Appendix) we estimate the same model with a two-way fixed effects specification, including both provider and country of residence fixed effects. This choice was made to consider individual-specific effects for people coming from the same country, which could be correlated with experienced utility. Results are robust and fully conform to the ones of the main models reported in Table 5.

In Table A5 (Appendix), we extend the investigation to include alternative specifications of the model and categorizations to identify the surprise effect. In Column 1, we explore the heterogeneity of the surprise effect (measured using PositiveSurprise and NegativeSurprise) across the two cities by introducing interactions between the surprise variables and the city dummy. We find a stronger positive impact of unexpected good weather in Milan compared to Venice, potentially attributed to the generally lower frequency of sunshine in Milan. Regarding negative news, Milan seems to experience a less severe negative impact from unexpected rain.[14] In Columns 2–4, we introduce alternative operationalizations of Surprise, namely, the difference between forecast and actual hours of sunshine (DiffHourSun) in Column 2, the disparity between predicted and actual rain probability (DiffRainProb) in Column 3, a dummy variable (SunnierIcon) that takes a value of 1 when the weather is sunnier than expected in Column 4 (similar to what is done in Table 5, Column 8, but this time using a dichotomous variable and running the model on the full sample rather than just on no surprise days). Again, the results are robust and consistent with the main model specification.

[13] We acknowledge the limitations of this variable, as it represents a linear shift from one status to another, whereas real-world effects may be non-linear.

[14] In this regard, one might think that results are driven by a selection bias: as typically, customers with strong evaluations (both positive and negative) are keener to post reviews than customers with mild feedback, in the dataset, there might be an over-representation of reviews in the days of strong (full) weather surprise. Attentive analysis shows that the share of reviews across the different states of the world (sunny, cloudy, and rainy days) is not statistically different from the share of realized weather in both cities. Similarly, the share of reviews reflects the share of days with positive, negative, or no surprises (results are available upon request). Hence, we can conclude that results are not driven by over- or under-representation of reviews in specific states of the world.

6. Concluding remarks

User-generated content has become a significant source of information that can help reduce the information asymmetry between buyers and service providers, especially in experience goods markets, where consumers face greater uncertainty. Subjects who provide feedback on their consumption experiences on rating platforms produce useful and reliable information which serves as a basis for future buyers when making their purchasing decisions. However, it is important to recognize that the experienced utility derived from consuming a service and shared through social media and online review systems is influenced by multiple factors beyond the service's intrinsic quality, such as prior expectations, situational factors, and unexpected circumstances.

Existing literature has shed light on some limitations and points out that reviews may come from a self-selected sample of users, who are either very happy or very unhappy, and may not necessarily represent a reliable metric of quality (Acemoglu et al., 2022; Li and Hitt, 2008; Hu et al., 2017). In this line of thinking, we posit that reviews may capture external elements and situational factors that are not directly under the control of the product or service providers, which could raise further concerns about the power of such informative content. We argue that the accommodation industry is an ideal context to test the relevance of unrelated and unexpected factors, given that the stay represents a typical example of experience good. Weather conditions are a key element that can significantly impact the enjoyment of a trip or a short holiday, being at the same time exogenous to the intrinsic quality of the hotel service to be evaluated and the entrepreneurial effort of the provider. Although several papers have already explored the effect of weather conditions on the reported utility of accommodation stays (Brandes and Dover, 2022; Jeuring, 2017; Gössling et al., 2016), we claim that the information set of consumers before travelling has not been taken into full account. Specifically, any mismatch between expectations and experience related to a situational factor like the weather has never been investigated.

In this paper, we fill this gap by defining a theoretical setting where agents form a belief about the weather conditions that will occur when travelling based on weather forecasts. These forecasts are a classic example of an informational public good, which is available to all agents at virtually no cost (Figini et al., 2022). A mismatch between the forecast and the realized weather, which we define as a *surprise*, representing a significant discrepancy between the expectation and realization of a particular event, could affect the experienced utility and be captured by online reviews. Therefore, holding the hotel quality constant, the experienced utility derived from consuming the accommodation service could differ according to how subjects are impacted by an exogenous shock, such as a sudden change in weather. The theoretical proposal of a 'surprise effect' on individual utility is not new and was discussed in Ely et al., 2015 and empirically tested in the context of entertainment and sports activities (Bryant et al., 1994; Peterson and Raney, 2008; Simonov et al., 2022; Bizzozero et al., 2016). However, our approach differs from previous studies in that we focus on the role of surprises coming as an exogenous shock rather than a fundamental aspect of the service being assessed.

We empirically test the model using individual ratings posted on Booking.com by customers who stayed in accommodation services in Milan and Venice between September 2019 and February 2020. The choice of both the review platform and the period under investigation is key to precisely identifying the alleged impact, limiting the importance of any confounding effect (e.g. non-verified reviews). Findings support the theoretical framework and show that: (i) the reported utility is affected by situational factors, namely the weather; (ii) differently from existing literature, we find that the driver is not only the weather *per se*, but also the surprise (the mismatch between expectations, proxied by weather forecasts, and the realized weather); (iii) a positive surprise (the realized weather is sunnier than expected) is associated with higher ratings, while a negative surprise (the realized weather is rainier than expected) is associated with lower ratings; (iv) consistently with good-news bad-news theory, we find evidence of negativity bias (Schwager and Rothermund, 2013), since positive surprises stemming from previous bad news about the weather exert stronger effects than negative surprises; (v) the effect is moderated by the length of stay, consistently with hedonic adaptation theory (Rayo and Becker, 2007; Frederick and Loewenstein, 1999).

We trust that the paper's findings do not stem from alternative mechanisms herein not considered. Some authors explain fluctuations in reported satisfaction levels in terms of projection bias, which refers to the tendency of individuals to overestimate the degree to which their future preferences will align with their current ones (Loewenstein et al., 2003). This cognitive bias could potentially affect the tourism context, given the temporal gap between the purchase decision and the actual service consumption. However, the projection bias involves the time of purchase, which is not accounted for in the current study because we lack data on when subjects have booked. Consequently, the effect of projection bias should not apply in this analysis.

To the best of the authors' knowledge, ours is a novel contribution to both the economics literature on the effect of surprises and, more in general, non-fundamental shocks on individual utility and the one on online reviews. First, we offer an empirical application of the effect of surprises on utility in a context that differs from most of the previous studies, which were mainly set in sports and media domains. Second, as per the literature on the effect of weather on recreational activities, our work investigates another facet of the relationship between weather-related information and agents' behaviour. Third, we provide robust evidence that online reviews might be affected by an additional bias, as they capture exogenous shocks that, to future customers, are irrelevant information. Such a bias might have non-neutral effects on subsequent reviews and future buying behaviours (Cicognani et al., 2021; Chevalier and Mayzlin, 2006).

As for practical takeaways, service providers need to be aware that they cannot control external factors affecting the quality, such as weather conditions. However, they can investigate how such effects can impact the utility and how to provide customers with alternative ways of enjoying their stay, even last minute if the weather turns out to be surprising. Furthermore, considering the climate crisis and the change in weather conditions, their predictability, and the increase in the frequency and intensity of extreme events, it is likely that service providers' operations and customers' evaluations will be impacted even more importantly in the future.

Finally, we also stress the relevance of weather forecast providers for the travel industry (Figini et al., 2022; Angelini et al., 2023), with non-neutral implications on consumers' enjoyment and, in turn, future behaviours. While forecast providers are working to have more accurate, reliable, and user-friendly weather forecasts, it is known that they apply a rain bias, given the asymmetric effect that a wrong forecast has on customers (Silver, 2012). Interestingly, such *wet bias* has two opposite effects on hotels: on the one hand, it might discourage people from booking a stay, given that they expect worse-than-optimal weather conditions, with a negative impact on revenues and prices; on the other hand, it might also trigger better rating scores, given the positive surprise effect found in this work when a forecast of rain translates in a sunny day.

We believe there is room for further research in this line of investigation, specifically in extending our methodology to different areas characterized by different external events, non-fundamental shocks, or different tastes in weather conditions. Additionally, analysing longer episodes, potentially characterized by multiple surprises, would allow exploring the extent to which the timing of the surprise plays a role in shaping utility. Moreover, having access to data on the purchase time, such as the distance between the booking date and consumption (the so-called 'lead time'), might help to control for the extent to which projection bias occurs. Specifically, it would enable the investigation of whether longer lead times exacerbate the misalignment between expected and actual satisfaction, thereby offering a more nuanced view of how this cognitive bias impacts consumer decision-making and subsequent satisfaction in the context of travel services. Finally, although our current understanding posits that weather surprises operate as unconscious factors, leading us to expect minimal mentions in textual reviews, a content and sentiment analysis of textual reviews could be a promising avenue to shed further light on the proposed mechanism at play.

Declaration of competing interest

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

Data availability

The data that has been used is confidential.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.jebo.2024.06.026.

References

Acemoglu, D., Makhdoumi, A., Malekian, A., Ozdaglar, A., 2022. Learning from reviews: The selection effect and the speed of learning. Econometrica 90 (6), 2857–2899.

Angelini, F., Figini, P., Leoni, V., 2023. High tide, low price? Flooding alerts and hotel prices in venice. Tour. Econ. 13548166231174270.

Baccan, D., Sbruzzi, E., Macedo, L., 2015. On the cognitive surprise in risk management: An analysis of the value-at-risk (VaR) historical. In: Progress in Artificial Intelligence. Springer International Publishing, pp. 402–413.

Bizzozero, P., Flepp, R., Franck, E., 2016. The importance of suspense and surprise in entertainment demand: Evidence from Wimbledon. J. Econ. Behav. Organ. 130, 47–63.

Brandes, L., Dover, Y., 2022. Offline context affects online reviews: The effect of post-consumption weather. J. Consum. Res. 49 (4), 595–615.

Bryant, J., Rockwell, S.C., Owens, J.W., 1994. "Buzzer Bearters" and "Barn Burners": The effects on enjoyment of watching the game go "down the wire". J. Sport Soc. Issues 18 (4), 326–339.

Bujisic, M., Bogicevic, V., Parsa, H., Jovanovic, V., Sukhu, A., 2019. It's raining complaints! how weather factors drive consumer comments and word-of-mouth. J. Hosp. Tour. Res. 43 (5), 656–681.

Busse, M.R., Pope, D.G., Pope, J.C., Silva-Risso, J., 2015. The psychological effect of weather on car purchases. Q. J. Econ. 130 (1), 371–414.

Camerer, C., Babcock, L., Loewenstein, G., Thaler, R., 1997. Labor supply of New York City cabdrivers: One day at a time. Q. J. Econ. 112 (2), 407–441.

Chevalier, J.A., Mayzlin, D., 2006. The effect of word of mouth on sales: Online book reviews. J. Mar. Res. 43 (3), 345–354.

Cicognani, S., Figini, P., Magnani, M., 2021. Social influence bias in ratings: A field experiment in the hospitality sector. Tour. Econ. 28 (8), 2197–2218.

Conlin, M., O'Donoghue, T., Vogelsang, T.J., 2007. Projection bias in catalog orders. Amer. Econ. Rev. 97 (4), 1217–1249.

Connolly, M., 2008. Here comes the rain again: Weather and the intertemporal substitution of leisure. J. Labor Econ. 26 (1), 73–100.

Cordell, V.V., 1997. Consumer knowledge measures as predictors in product evaluation. Psychol. Market. 14 (3), 241–260.

Dehaan, E., Madsen, J., d. Piotroski, J., 2017. Do weather-induced moods affect the processing of earnings news? J. Account. Res. 55 (3), 509–550.

DeMarzo, P.M., Vayanos, D., Zwiebel, J., 2003. Persuasion bias, social influence, and unidimensional opinions. Q. J. Econ. 118 (3), 909–968.

Deng, L., Sun, W., Xu, D., Ye, Q., 2021. Impact of anonymity on consumers' online reviews. Psychol. Market. 38 (12), 2259–2270.

Derbaix, C., Vanhamme, J., 2003. Inducing word-of-mouth by eliciting surprise – a pilot investigation. J. Econ. Psychol. 24 (1), 99–116.

Easterlin, R.A., 1995. Will raising the incomes of all increase the happiness of all? J. Econ. Behav. Organ. 27 (1), 35–47.

Eil, D., Rao, J.M., 2011. The good news-bad news effect: Asymmetric processing of objective information about yourself. Am. Econ. J. Microecon. 3 (2), 114–138.

Ely, J., Frankel, A., Kamenica, E., 2015. Suspense and surprise. J. Polit. Econ. 123 (1), 215–260.

Figini, P., Cicognani, S., Zirulia, L., 2022. Booking in the rain. Testing the impact of public information on prices. Italian Econ. J. Online first: http://dx.doi.org/10.1007/s40797-022-00199-y.

Figini, P., Vici, L., Viglia, G., 2020. A comparison of hotel ratings between verified and non-verified online review platforms. Int. J. Cult. Tour. Hosp. Res. 14 (2), 157–171.

Fradkin, A., Grewal, E., Holtz, D., 2018. The determinants of online review informativeness: Evidence from field experiments on airbnb. SSRN Electron. J..

Frederick, S., Loewenstein, G., 1999. 16 Hedonic adaptation. In: Kahneman ED, D., Schwarz, N. (Eds.), Well-Being the Foundations of Hedonic Psychology. pp. 302–329.

Gössling, S., Abegg, B., Steiger, R., 2016. "It was raining all the time!": Ex post tourist weather perceptions. Atmosphere 7 (1), 10.

Greiff, M., Paetzel, F., 2020. Information about average evaluations spurs cooperation: An experiment on noisy reputation systems. J. Econ. Behav. Organ. 180, 334–356, URL https://www.sciencedirect.com/science/article/pii/S0167268120303826.

Helmers, C., Krishnan, P., Patnam, M., 2019. Attention and saliency on the internet: Evidence from an online recommendation system. J. Econ. Behav. Organ. 161, 216–242, URL https://www.sciencedirect.com/science/article/pii/S0167268119301131.

Hu, N., Pavlou, P.A., Zhang, J., 2017. On self-selection biases in online product reviews. MIS Q. 41 (2), 449–475.

Huysmans, F., 2002. Influence of the weather on time use.

Ifrach, B., Maglaras, C., Scarsini, M., Zseleva, A., 2019. Bayesian social learning from consumer reviews. Oper. Res. 67 (5), 1209–1221.

Janakiraman, N., Meyer, R.J., Morales, A.C., 2006. Spillover effects: How consumers respond to unexpected changes in price and quality. J. Consum. Res. 33 (3), 361–369.

Jeuring, J.H.G., 2017. Weather perceptions, holiday satisfaction and perceived attractiveness of domestic vacationing in The Netherlands. Tour. Manag. 61, 70–81.

Jeuring, J.H., Peters, K.B., 2013. The influence of the weather on tourist experiences. J. Vacat. Market. 19 (3), 209–219.

Kahneman, D., Thaler, R., 1991. Economic analysis and the psychology of utility: Applications to compensation policy. Am. Econ. Rev. 81 (2), 341–346.

Kahneman, D., Thaler, R.H., 2006. Anomalies: Utility maximization and experienced utility. J. Econ. Perspect. 20 (1), 221–234.

Kettlewell, N., Morris, R.W., Ho, N., Cobb-Clark, D.A., Cripps, S., Glozier, N., 2020. The differential impact of major life events on cognitive and affective wellbeing. SSM - Popul. Health 10, 100533.

Kim, J.M., Han, J., 2022. Impact of the length of stay at hotels on online reviews. Int. J. Contemp. Hosp. Manag. 34 (4), 1249–1269.

Klimstra, T.A., Frijns, T., Keijsers, L., Denissen, J.J.A., Raaijmakers, Q.A.W., van Aken, M.A.G., Koot, H.M., van Lier, P.A.C., Meeus, W.H.J., 2011. Come rain or come shine: Individual differences in how weather affects mood. Emotion 11 (6), 1495–1499.

Lee, J.J., Gino, F., Staats, B.R., 2014. Rainmakers: Why bad weather means good productivity. J. Appl. Psychol. 99 (3), 504–513.

Leoni, V., Boto-García, D., 2023. 'Apparent'and actual hotel scores under booking. com new reviewing system. Int. J. Hosp. Manag. 111, 103493.

Li, X., Hitt, L.M., 2008. Self-selection and information role of online product reviews. Inf. Syst. Res. 19 (4), 456–474.

Loewenstein, G., O'Donoghue, T., Rabin, M., 2003. Projection bias in predicting future utility. Q. J. Econ. 1209–1248.

Manski, C.F., 1993. Identification of endogenous social effects: The reflection problem. Rev. Econ. Stud. 60 (3), 531.

Mayzlin, D., Dover, Y., Chevalier, J., 2014. Promotional reviews: An empirical investigation of online review manipulation. Amer. Econ. Rev. 104 (8), 2421–2455.

Nelson, P., 1970. Information and consumer behavior. J. Polit. Econ. 78 (2), 311–329.

Nguyen, V.H., Claus, E., 2013. Good news, bad news, consumer sentiment and consumption behavior. J. Econ. Psychol. 39, 426–438.

Oswald, A.J., Powdthavee, N., 2008. Does happiness adapt? A longitudinal study of disability with implications for economists and judges. J. Public Econ. 92 (5–6), 1061–1077.

Peterson, E.M., Raney, A.A., 2008. Reconceptualizing and reexamining suspense as a predictor of mediated sports enjoyment. J. Broadcast. Electron. Media 52 (4), 544–562.

Rayo, L., Becker, G.S., 2007. Evolutionary efficiency and happiness. J. Polit. Econ. 115 (2), 302–337.

Reimers, I., Waldfogel, J., 2021. Digitization and pre-purchase information: the causal and welfare impacts of reviews and crowd ratings. Amer. Econ. Rev. 111 (6), 1944–1971.

Riis, J., Loewenstein, G., Baron, J., Jepson, C., Fagerlin, A., Ubel, P.A., 2005. Ignorance of hedonic adaptation to hemodialysis: A study using ecological momentary assessment. J. Exp. Psychol.: Gen. 134 (1), 3–9.

Schoenmueller, V., Netzer, O., Stahl, F., 2020. The polarity of online reviews: Prevalence, drivers and implications. J. Mar. Res. 57 (5), 853–877.

Schützwohl, A., 1998. Surprise and schema strength. J. Exp. Psychol: Learn. Mem. Cogn. 24 (5), 1182–1199.

Schwager, S., Rothermund, K., 2013. Motivation and affective processing biases in risky decision making: A counter-regulation account. J. Econ. Psychol. 38, 111–126.

Silver, N., 2012. The Signal and the Noise: Why So Many Predictions Fail-But Some Don't. Penguin.

Simonov, A., Ursu, R.M., Zheng, C., 2022. Suspense and surprise in media product design: Evidence from twitch. J. Mar. Res. 60 (1), 1–24.

Sims, C.A., 2006. Rational inattention: Beyond the linear-quadratic case. Amer. Econ. Rev. 96 (2), 158–163.

Soroka, S.N., 2006. Good news and bad news: Asymmetric responses to economic information. J. Politics 68 (2), 372–385.

Sunder, S., Kim, K.H., Yorkston, E.A., 2019. What drives herding behavior in online ratings? The role of rater experience, product portfolio, and diverging opinions. J. Market. 83 (6), 93–112.

Wilson, T.D., Gilbert, D.T., 2008. Explaining away: A model of affective adaptation. Perspect. Psychol. Sci. 3 (5), 370–386.