CrossMark

ORIGINAL RESEARCH

The impact of the package opening process on product returns

Wenyan Zhou[1] · Oliver Hinz[2] · Alexander Benlian[1]

Abstract High product return rates are an increasingly pressing challenge for many e-retailers around the world. To address this problem, this paper offers a new perspective by focusing on the critical moment of the package-opening process. Going beyond previous research, which has primarily focused on website information and the product itself, we examine the effects of the outside appearance (i.e., the color of the delivery package) and contents of the delivery package (i.e., extra gifts, coupons, and preprinted return labels) on consumer return behavior. Our findings across two experimental studies and an observational field study show that a well-considered package design, including colorful packaging and extra gifts, significantly lowers consumers' return intentions and actual returns. We also explore the process of consumers' cognitive–affective reactions after opening a delivery package. During this two-stage reaction process, pleasure plays a crucial role in the consumer's return choice.

✉ Wenyan Zhou
    wenyan.zhou@stud.tu-darmstadt.de

    Oliver Hinz
    ohinz@wiwi.uni-frankfurt.de

    Alexander Benlian
    benlian@ise.tu-darmstadt.de

[1] Faculty of Law and Business, Technische Universität Darmstadt, Germany, Hochschulstraße 1, 64289 Darmstadt, Hessen, Germany

[2] Chair of Information Systems and Information Management, Faculty of Economics and Business Administration, Goethe University Frankfurt, Theodor-W.-Adorno-Platz 4, 60323 Frankfurt am Main, Germany

🎄 Springer

1 Introduction

Rising online return rates pose a serious threat to e-retailers around the world. One recent survey showed that 36.1% of German e-retailers selling fashion and accessories experienced a return rate of 20% or more in 2014 (Institut für Demoskopie Köln 2015). In the United States, the average return rate reached 33% in 2012 and is projected to increase even further in the future (Banjo 2013). To compound matters, 72% of e-retailers bear the costs of delivery and/or the extra labor and management costs for handling the returns, which can ultimately render a lenient return policy very expensive. Researchers and managers are currently seeking ways to mitigate this problem, but our understanding of what drives high return rates remains limited.

Researchers agree that online purchasing can be viewed as a two-stage decision process (Wood 2001): the decision to order (the purchase decision) and the decision to keep or return the ordered product (the post-purchase decision). The purchase decision itself is more time-consuming than the post-purchase decision: Consumers typically spend a great deal of time collecting and processing information from several e-retailers and comparing different products, brands, and prices. In contrast, they usually make the post-purchase decision quickly, sometimes within a few seconds after opening the package. Therefore, it is critical to determine the factors that influence the post-purchase decision during this brief yet decisive period.

A vast array of literature on consumer return behavior pertains to website information and the product itself (see Fig. 1), primarily using signaling (Janaki-raman and Ordóñez 2012), utility (Wood 2001), and expectation (Bechwati and Siegal 2005) theories. In general, research indicates that a good reputation (Zhou and Hinz 2016), a high review score (Sahoo et al. 2015), and higher-quality products (Anderson et al. 2009) can reduce consumer return intentions by increasing perceived utility and signaling higher quality. Other studies show that a lenient return policy can increase returns because of lower return costs for the consumer (Wood 2001), while an accurate product description on purchase websites can decrease return intentions by narrowing consumers' expectation gap (Heiman et al. 2001).

However, because of the time delay between purchasing and receiving a product ordered online, consumers may not clearly remember all the purchase details at the moment of receipt and may not visit the respective website or check old e-mails to reacquaint themselves with actual purchase information or return policies. As a result, above and beyond purchase details, delivery package cues are likely one of the last opportunities for e-retailers to influence consumers' product return behavior (Garretson and Burton 2005). The information carried by the delivery package is probably clearer and fresher than what remains in consumers' memories (i.e., the stimuli shown at the moment of purchase). Thus, to address the problem of high return rates, both e-retailers and researchers need to recognize the importance of delivery package design, even though little, if any, extant research has done so.
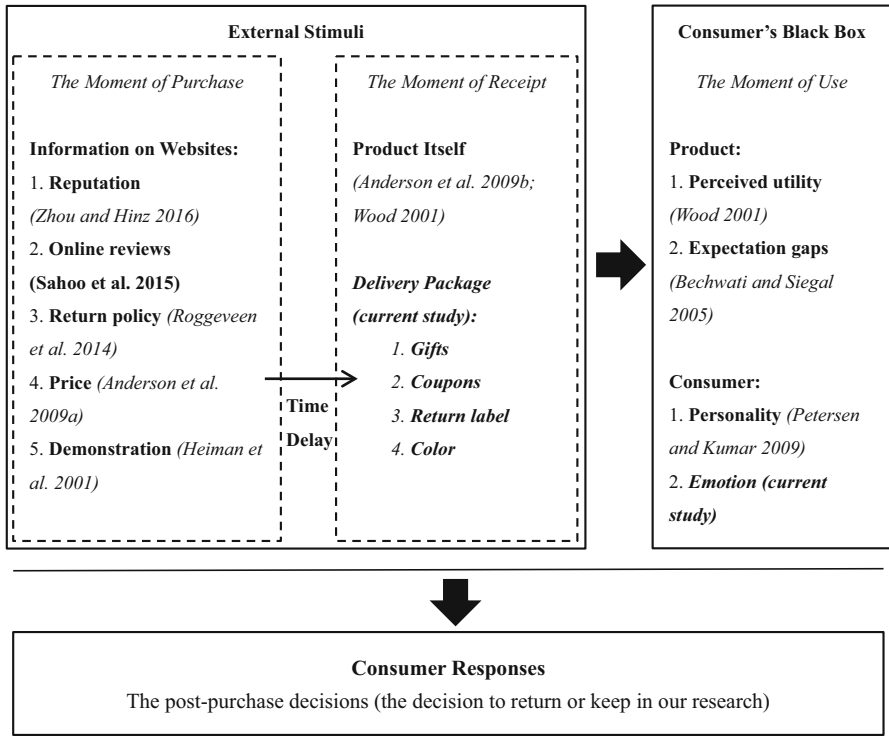
Fig. 1 Factors that can influence the post-purchase decisions in different moments

The composition of the delivery package design includes the outside appearance (e.g., color, shape, size) and its contents (e.g., extra gifts, coupons, return labels, fillers, and receipts). We focus on color, extra gifts, coupons, and return labels as the four most critical aspects in our study, because they do not heavily rely on the characteristics of ordered products (i.e., we do not consider shape, size, and fillers) and can potentially influence consumer behavior (see Sect. 2). Moreover, e-retailers can easily manipulate all these factors.

Against this background, we pose the following research question: How does the package-opening process influence consumers' return decision and why? The question includes two parts: consumers' responses to external stimuli (i.e., the delivery package design) at the moment of opening the package and their reaction process when they make return decisions. Answering these questions is crucial for research and practice alike. Theoretically, the study advances research on the drivers of consumer return behavior and sheds more light on the return decision process. A new perspective on the package-opening process would also aid in analyzing other consumer post-purchase decisions. Practically, addressing this research question identifies several controllable factors that e-retailers can alter to reduce return rates.

This article proceeds as follows: In Sect. 2, we review the literature on consumer return behavior, as well as the potential impact that designed packages can have on consumer behavior. In Sect. 3, we present an experiment (wherein we imitate the

purchasing and package-opening process using stop-motion videos) that tests the impact of package design on consumers' return intentions (Study 1). We further apply partial least squares (PLS) regressions to understand consumers' cognitive and affective reactions during the opening process. Section 4 describes Study 2, which serves as a robustness test and extends the analysis to real return behavior in an experimental setting. In Sect. 5, we use transactional data from China's largest online platform (Taobao.com) to examine the relationship between package design and a web shop's return rates (Study 3). Section 6 concludes with a discussion of the implications and future research avenues.

2 Literature review and conceptual framework

Our study aims to clarify the entire process between consumers' reaction during the package opening period to their final decision to return or keep. Thus, we consider three types of processes that occur during one's exposure to a stimulus event (Berkowitz 1993; LeDoux 1995; Shiv and Fedorikhin 1999), which is the opening of a delivery package. The first is the 'low-road' affective[1] processes, which are evoked rapidly and automatically in the limbic systems of the brain. The second is the 'high-road' cognitive processes that involve the cortical systems of the brain. In this stage, consumers collect and process information by thinking, reasoning, and being aware of their surroundings. The third is the 'high-road' affective processes, which arise from the outcomes of the second process (Shiv and Fedorikhin 1999), and occur relatively slowly compared to low-road affective reactions. All these affective and congnitive reactions can influence consumer return behavior and will be discussed in the next sections.

2.1 Cognitive reactions in consumer return behavior

Prior literature usually considers consumer return behavior a consequence of cognitive reactions. According to utility theory, consumers return products only if the net perceived utilities (i.e., the utilitarian utility and hedonic utility) are negative. The net perceived utilities are evaluated by a customer's perceived utilities at the purchase stage, the perceived fit of the transaction (e.g., physical fit or sensory related to color), and the return costs (Anderson et al. 2009) at the moment of the receipt. In the situation of online shopping, the perceived utilities (including utilitarian and hedonic utilities) are generated by product information given on websites and stored in consumers' memories. The perceived utilities can be increased by lower purchase prices (due to a higher customer surplus) (Anderson et al. 2009a), a good reputation, and a lenient return policy [i.e., longer deadlines, less required return effort, and more coverage (Roggeveen et al. 2014)]. Return

---

[1] The 'low-road' affective process is highly evolutionary and is designed to protect individuals from life-threatening danger, and to elicit defensive responses without conscious thought. The 'high-road' affective/cognitive processes, by contrast, involve an indirect pathway to the amygdala. In this case, thalamic info is transmitted to the sensory cortex where it is further processed and evaluated for level of threat prior to being sent to the amygdala LeDoux (1995).

costs are the fees that consumers must bear if they decide to return the ordered products, such as restocking fees and frictional costs for bringing the parcel to a post office (Frischmann et al. 2012), which could be reduced by a lenient return policy (e.g., a preprinted return label).

The expectation gap (also called fit of transaction in some articles) is a term frequently mentioned in the consumer return behavior literature (Anderson et al. 2009; Petersen and Kumar 2009). An expectation gap exists when the perceived utilitarian and hedonic utilities at the moment of receipt are not equal to those perceived by the consumer at the moment of purchase. This phenomenon can be attributed to the following two characteristics of the two-stage decision process: (1) the time delay between payment and receipt, which allows disconfirming information (e.g., advertisements from competitors) to provide a potential negative influence (Bechwati and Siegal 2005); and (2) the information asymmetry between sellers and buyers (Anderson et al. 2009), which can result in a difference between the expectations created by website information and the perceived utility of the actual product. Larger negative expectation gaps at the moment of product receipt can increase consumers' return intentions. There are basically two solutions to narrow the negative expectation gap: One way is calibrating consumers' perceived utility at the moment of purchase. E-retailers can create reliable expectations by offering more detailed and accurate descriptions, as well as services like user-generated product evaluations (Zhou and Hinz 2016). Alternatively, e-retailers can enhance consumers' perceived utilities at the moment of receipt, e.g., through a well-designed package.

Although prior research has paid adequate attention to the cognitive reactions in consumers' return behavior, no study has, to our knowledge, considered the potential impact of the delivery package. A delivery package with a suitable color, extra gifts, or coupons might also enhance the perceived utilities of the entire transaction, which could ultimately decrease consumers' return intentions and actual return behavior.

2.1.1 Color and cognitive reactions

Color is an aesthetic stimulus widely used in marketing (i.e., advertisement, logo, brand personality, package, and atmospherics) to grab consumers' attention, enhance purchase intention, and increase perceived service quality (Labrecque et al. 2013; Puccinelli et al. 2013). Color can convey both embodied meaning (discussed in Sect. 2.2) and referential meaning (Zeltner 1975). Referential meaning emerges from the network of semantic associations and, as such, is a dynamic and reflexive process (Elliot and Maier 2007). According to memory models, people store semantic information in a complex network comprising conceptual nodes (external stimuli) and links (pathways between nodes) (McClelland 1988). As nodes become activated (e.g., through stimulation by colors), the activation spreads to additional nodes through links. For instance, many premium brands (e.g., IBM, Walmart, Volkswagen) use blue in their logos and product package designs because consumers link blue with competence; the color is further associated with intelligence, communication, trust, efficiency, duty, and logic. Thus, a brand with blue hues can positively affect consumers' perceptions of the brand's competence

(Labrecque and Milne 2012), which can lead to greater perceived utilities. However, these associations can differ among various cultures (Madden et al. 2000). For instance, the color "red" in China signals auspiciousness, while in the western world, "red" is related to anger (Jacobs et al. 1991).

The delivery package design can also serve as part of a brand image. Thus, suitably colored delivery packages can activate the node of a competent brand and/ or high-quality products, and thereby increase consumers' evaluations of the entire transaction.

2.1.2 Gift, coupons, and cognitive reactions

The two main categories of sales promotions are nonmonetary (e.g., extra gifts) and monetary (e.g., coupons), which provide consumers with an array of hedonic and utilitarian benefits (Chandon et al. 2000). Specifically, the hedonic benefits include value expression (i.e., the expression and enhancement of the self-concept and personal values) entertainment, and exploration, while utilitarian benefits include savings, quality, convenience, and value expression. While many aspects of Chandon et al. 2000 "benefit congruency framework" do not apply to the post-purchase stage, typical sales promotions can still exert positive influences (Liu and Chou 2015). Suitable extra gifts can increase utilitarian benefits by saving money that would have gone to additional products (e.g., accessories to the products purchased) or by providing a convenient way to try new products (e.g., a trial product). Moreover, extra gifts can create hedonic benefits by offering entertainment functions (e.g., a whistle as a gift with a football jersey).

Coupons bring more utilitarian benefits than hedonic ones (Chandon et al. 2000). Specifically, coupons could result in money savings by offering discounts to be redeemed with the next purchase.

In sum, extra gifts and coupons in delivery packages may increase perceived utilities by creating additional utilitarian and hedonic experiences, which can ultimately reduce product returns.

2.2 Affective reactions in consumer return behavior

Consumers' affective reactions have two main dimensions: pleasure and arousal (PA model) (Chang et al. 2014; Ladhari 2007; Mazaheri et al. 2014; Mazaheri et al. 2012). Pleasure is the degree to which consumers feel happy, good, contented, or joyful (Mehrabian and Russell 1974); arousal is the degree to which consumers feel excited, stimulated, awake, or active (Mehrabian and Russell 1974). Both pleasure and arousal are important affective responses and can be influenced by a well-considered package design. In contrast to the widespread concern about cognitive reactions, research attention to the affective reactions (e.g., emotions) in consumer return behavior has been limited to date.

We found no literature referring to the multiple-stage reaction process in a product return context—not even studies solely addressing affective reactions. The only relevant stream of literature we found is situated in the context of service failures (Aurier and Guintcheva 2014; Ladhari 2007; Oliver 1993). Service failures

and product returns entail a similar decision-making process, insofar as affected consumers can apply for a refund or compensation when they are not satisfied with the offered service/product. In the context of service failures, the type and degree of a service failure, in tandem with the offered remedial measures, can sway consumers' emotions. In line with the Appraisal-Tendency framework (Scherer et al. 2001), these affective responses can then influence consumers' post-purchase decisions (e.g., loss or maintenance of access to goods/services) and ultimately their satisfaction (Hibbert et al. 2012) and loyalty (DeWitt et al. 2008). Specifically, when consumers have more negative emotions towards a service failure or are more aware of it, they are less likely to be satisfied and might not maintain their purchase decisions (Choi and Mattila 2008).

It is worth noting that the emotions mentioned in previous service failures research are "high-road" affective reactions (i.e., anger and regret), as they are the outcomes of cognitive reactions (i.e., appraisals about service failure) (Bonifield and Cole 2007; Choi and Mattila 2008; Gelbrich 2010; Smith and Bolton 2002). However, marketing researchers have also explored "low-road" affective reactions, e.g., emotions evoked by color, gift, and coupon. As we mentioned before (in Sect. 2.1.1), colors convey an embodied meaning, which is not learned but rather driven by stimulation that is embodied in colors (Meyers-Levy and Peracchio 1995). Thus, marketers commonly use a long-wavelength color (e.g., red) in the pre-purchase stage (e.g., advertisement) to stimulate arousal (Mehta and Zhu 2009) (e.g., excitement) and tend to use a short-wavelength color (e.g., blue) in the post-purchase stage to lower arousal (e.g., relaxation). Gifts or coupons, as an effective promotion strategy, can create a positive surprise when they appear in the package (Heilman et al. 2002). This positive surprise may cause consumers to experience greater perceived pleasure and perhaps even arousal.

However, as the multiple-stage reaction process for return or keep decisions passes quickly, it is difficult to clearly distinguish "low-road" affective reactions from "high-road" ones. Nonetheless, given that return behavior requires deliberation, we decided to focus on a two-stage "high-road" cognitive–affective reaction process, in line with other studies on service failure (Chitturi et al. 2007; Schoefer 2010).

Finally, researchers have also identified personality (Petersen and Kumar 2009), demographic variables, and industry characteristics (Anderson et al. 2009) as influential factors in consumer return behavior; therefore, we controlled for these factors. Figure 2 shows the resulting conceptual framework of our study.

We conducted three empirical studies to estimate the impact of various package designs on consumer return behavior (see Fig. 2). Studies 1 and 2 are controlled experiments. Study 1 focuses on the impact of the package-opening process on consumers' return intentions, while Study 2 observes consumers' real return behavior after receiving packages in different designs. Study 3 is a field study, exploring the relationship between package designs and e-retailers' return rates in a real online market. In both studies 1 and 2, we used blue as the hue for a colorful package and brown for the control group. In Study 3, we compared the commonly used color for packages (i.e., brown) to all the other colorful hues (e.g., red, blue and black). The consistent results of all three studies provide convincing evidence that the package-opening process can influence consumers' return decisions.
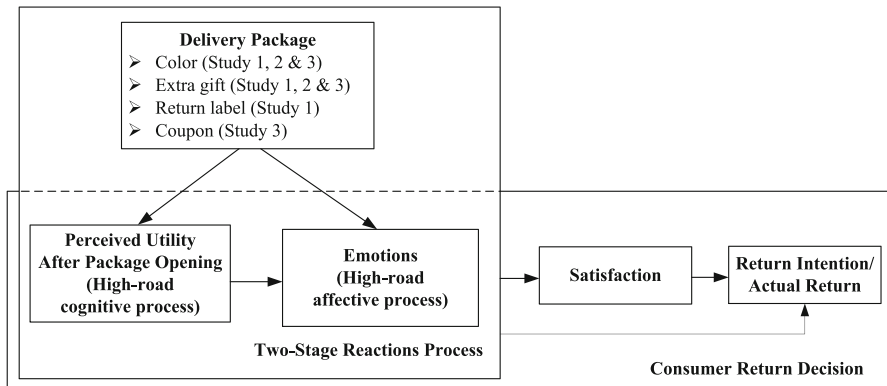
Fig. 2 Conceptual framework

3 Study 1: An experiment on the package-opening process with soccer jerseys

In Study 1, we designed an experiment with eight different treatments and invited 375 (43 for a pretest and 332 for the main experiment) subjects to participate. We first used analysis of variance (ANOVA) and analysis of covariance (ANCOVA) to examine the influence of colorful packages, extra gifts, and preprinted return labels on consumers' return intentions. We then estimated a PLS model to better understand how these effects operate through consumers' cognitive–affective reaction processes.

3.1 Methods

We based our choice of package stimuli on related research and the pretest results. For the pretest, we invited 43 German participants [50% female; average age
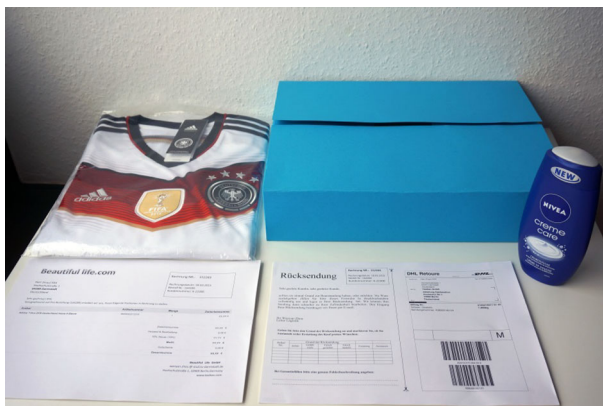
Fig. 3 Product and package stimuli in Study 1

36.1 years (SD = 10.20)] to provide feedback on different manipulations. Figure 3 shows the final treatments for both the product and package stimuli. The results of the pretest also show that our scales achieved good reliability.

3.1.1 Product

Four criteria guided our selection of a product: (1) a high proportion of our target population should be interested in buying this product, (2) the product should belong to an industry whose return rate is relatively high, (3) the product should have both utilitarian and hedonic value, and (4) a product defect can be easily manipulated to enhance consumer return intention after the package opening. Keeping all these criteria in mind, we selected a jersey of the German national football team and added a 5 mm × 5 mm black stain on the back. We decided to use only one black stain because in our pretest, as more than one mark led to extremely high return rates (> 90%), strongly reducing the variance in our dependent variable. To ensure that people noticed the stain and assessed the problem similarly, we stated that "You have no idea what caused the stain, but you notice that you might not be allowed to return the jersey after washing it."

3.1.2 Color

According to a report in 2001, 40% of Germans' favorite color is blue, followed by red (19%) and then green (18%) (Institut für Demoskopie Allensbach 2001). In our pretest, both men and women indicated that the color of an ideal delivery package, other than standard brown, was blue. Thus, we chose blue-colored delivery packages for our experiment. Crowley (1993) documents that blue has a strong impact on shopping in terms of both evaluation and activation, which meets the requirements of our research goal. The control group received a delivery package in standard brown.

3.1.3 Gift

In the pretest, we also tested the estimated price of various extra gifts. In line with the results, we selected Nivea Creme Care as the extra gift for our main test. The price (approximately €2.5) is 3% of the price of a soccer jersey, and both men and women can use it.

3.1.4 Preprinted return label

We placed a preprinted and prepaid DHL label with a return shipping address into the package. To return the package, participants needed only to glue this return label to the original delivery package and bring it to a post station. In the control group, participants needed to log into their accounts, complete several forms, and then print the return document themselves. We reasoned that a preprinted, prepaid DHL label could significantly reduce return costs and thus, according to utility theory, increase consumers' return intentions.

3.1.5 Design and procedure

We employed a 2 (colorful vs. not colorful) × 2 (gift vs. no gift) × 2 (preprinted return label vs. no preprinted return label) between-subjects design on the online survey platform Dynamic Intelligent Survey Engine.[2] In step 1, we randomly assigned participants to one of the eight experimental conditions and asked for demographic information (i.e., age, gender, and career). In step 2, we simulated an online purchase process. Participants were asked to imagine that they had decided to buy a jersey of the German national football team for the upcoming World Cup and then to specify their size and gender in order to obtain the appropriate jersey.

In step 3, we clarified that they were to imagine that they paid for their selected jersey, and then we asked for their emotions (pleasure and arousal) toward and perceived utility (utilitarian and hedonic) of the jersey. For step 4, we needed to create an artificial time delay between the payment and the virtual receipt. Thus, we employed a filler task in which participants answered questions about their online shopping experience and personality by identifying the extent of their extroversion, agreeableness, conscientiousness, neuroticism, and openness (a 10-item short version of the Big Five Inventory in German) (Rammstedt and John 2007). Afterward, participants learned that "after 3 days, you receive your order." Subsequently, in step 5, we told participants, "Please assume that you were the person who opened the package in the video" and then used a 30-s stop-motion animation to show the entire opening process. In stop-motion (also known as stop-frame) animation, an object (in this case, the package) is moved in small increments between individually photographed frames, creating the illusion of movement when the series of frames is played as a continuous sequence. This technique allowed us to control the timing and method of package opening. The eight videos in the eight experimental groups were exactly the same except for our manipulations. The gift and/or the preprinted return label appeared for approximately 5 s (six photos for the process taking the items from the package, two photos for a full-screen display of the details, and another two photos for putting the items down; for details, see Table 1). We used an amplification process for the gift and preprinted return label to ensure that every participant could recognize each stimulus clearly. Participants could not move to the next step until they finished watching the whole video.

In step 6, we surveyed participants' current emotions and the perceived utility of the whole package, along with their satisfaction and return intentions. To keep the package in participants' minds, we placed a picture of the package, showing all the items, at the top of the questionnaire (see Fig. 3). In step 7, in order to match their return intention to real return behavior, we communicated that every participant had a chance to win the real package shown in the video (with extra gift/colorful package/preprinted return label and a jersey with a stain) and that they could send the jersey back for a new, flawless one. We then asked whether they would really return their jersey in that case. As an additional motivation and to increase realism, we asked participants to voluntarily give their contact information and jersey size.

---

[2] The Dynamic Intelligent Survey Engine, or DISE, is a powerful survey engine built to support and facilitate a variety of advanced data collection methods. See http://www.dise-online.com/.

Table 1 Details of the stop-motion animation video

Manipulation	Number of photos	Display time (s)
Colorful package	60	25 s
Extra gift	10	5 s
Preprinted return label	10	5 s
Dirty-stained jersey	30	10 s (5 s for the dirty stain)
Total	80[a]	Approximately 30 s[a]

[a] Some photos include more than one stimulus (e.g., participants saw the extra gift in the colorful package)

In the final step 8, we randomly chose five participants and sent them the package exactly as shown in the video of their treatment group and asked them whether they would like to return the flawed jersey. If they wanted to return, they had to bring the parcel to the post office and had to wait until they received their flawless jersey. This additional step allowed us to observe their real return decision and examine whether their answers (return intentions) in the experiment matched their real behavior. Figure 4 summarizes the entire experimental procedure.

3.1.6 Scales

We adapted our items for measuring the constructs from prior marketing research (see Table 2) using multi-item Likert-type scales for each. We assessed perceived utility using the hedonic/utilitarian scale proposed by Voss et al. (2003). This scale includes eight-point semantic differential items, but we decided to use only seven points according to the Cronbach's α results ($> .7$). Moreover, we measured emotions using the PA model (including three items for pleasure and three items for arousal) from Mehrabian and Russell (1974). We measured perceived utility and emotion twice—once after participants' purchase decisions and again after the package-opening process. Note that the perceived utility tested following the package opening pertains to the whole package. For consumer satisfaction, we adopted Finn (2005) three-item scale, which is widely used in marketing research.

To assess consumer return intention, we used the Net Promoter Score (NPS), which is based on an 11-point Likert scale (0 = "not at all likely" and 10 = "very likely") introduced by Reichheld (2003) and widely used to measure attitudes or behavioral intentions (Samson 2006). The NPS is calculated with a single question, in our case, "How likely is it that you would return the package?" We identified participants who responded with a score of 9 or 10 on the NPS as package returners and those who responded with a score of 0–6 as package keepers.

In the real return behavior check (Step 7 and 8), we coded participants' answers with a dummy variable equal to 0 if they claimed to keep the whole package shown in the video and 1 if they opted to send it back to get a new one. Although receiving a gift is different from a real purchase, the return decision is similar in our simulated case. Thus, we believe participants' choice of gift return can proxy for their actual
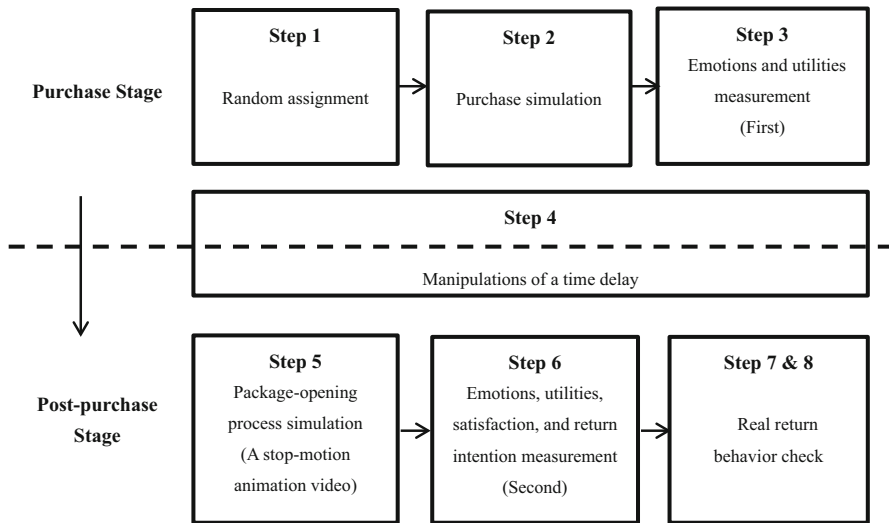
Fig. 4 Experimental procedure of Study 1

Table 2 Scale items

Construct	Item	Cronbach's α	CR	AVE
Utility				
Utilitarian utility	Ineffective/effective	.929	.949	.824
	Helpful/unhelpful			
	Functional/not functional			
	Practical/impractical			
Hedonic utility	Fun/not fun	.918	.948	.859
	Delightful/not delightful			
	Enjoyable/unenjoyable			
Emotion				
Pleasure	Happy/unhappy	.941	.962	.895
	Pleased/annoyed			
	Contented/melancholic			
Arousal	Relaxed/stimulated	.749	.846	.648
	Calm/excited			
	Dull/jittery			
Satisfaction				
	I feel comfortable with the package.	.943	.963	.898
	The package is satisfying to me.			
	The package is worth the time and money I spent on it.			

behavior after receiving a product with a small flaw. We then compared participants' return intention (0–6 for non-return, 9–10 for return) and their real return choice (0 for non-return, 1 for return); these two answers were highly correlated ($p < .01$).

All survey items were presented in the respondents' native language (German). We pretested the final questionnaire with doctoral students and university employees to identify unclear and ambiguous questions. The convergent and discriminant validity for the constructs exceeded all critical values (see Table 2).

3.2 Sample

After conducting a pretest with 43 participants who came from our target population of native Germans with Internet access, we employed a professional market research company to collect a representative sample for our main study in March 2015. Our initial sample for our main study included 332 participants, all of whom had recent online shopping experience. To keep our sample representative within each experimental group, we set quotas for age and gender according to Europe's 2014 online shopping consumer report (Eurostat 2014). To verify the validity of the responses, we checked each participant's response patterns and completion time. We excluded five questionnaires that were completed in less than five minutes, six questionnaires that exhibited a visible pattern of the same response on all the Likert scales, and one questionnaire from a participant who reported that his computer was unable to play the video. The final sample thus consisted of 320 completed surveys (see Table 3). An ANOVA revealed no significant differences in participants' age, gender, occupation, and soccer preference among the eight experimental groups, which indicates that our randomization worked as intended.

Table 3 Descriptive statistics

	Percentage (%)	SD
Age (years)		
16–24	22.5	10.21
25–34	31.8	
35–44	25.6	
45–54	20.0	
Number of online shopping experiences in the past 6 months		
0	.93	.77
1–3	22.50	
4–10	41.88	
> 10	34.69	
Individual return rate after online shopping		
0–5%	34.06	.84
5–10%	45.31	
10–20%	15.00	
> 20%	5.63	

3.3 Common method bias analysis

We strived to design the questionnaire carefully, which entailed ensuring participants' anonymity, using a random order for survey items, providing concrete survey instructions, and asking participants to answer the questions as honestly as possible (Podsakoff et al. 2003). Nevertheless, self-reported data can suffer from common method biases, such as consistency motifs or social desirability concerns (Podsakoff et al. 2003). Thus, we adopted the marker variable approach (Rönkkö and Ylitalo 2011) to test whether a common method bias confounded our results.

We performed the marker variable method (Rönkkö and Ylitalo 2011) with two marker items (two items for Openness, which the ANCOVA in Table 4 shows to be unrelated to the dependent variables) taken from our empirical data set; these items were not included in our research model and lack an explicit theoretical influence on the constructs in our research model. Following Rönkkö and Ylitalo's (2011) method, we found relatively low correlations between the marker items and study items (the mean values of the correlation coefficients were .046 and .061) and determined that these low correlations must have been caused by the method. Next, we included the marker items as additional latent variables in our PLS analysis

Table 4 Results of ANOVA and ANCOVA

Variables	ANOVA		ANCOVA	
	Satisfaction F value	Return intention F value	Satisfaction F value	Return intention F value
Color (C)	1.787	3.124*	1.764	3.660*
Gift (G)	9.607***	5.036**	10.685***	4.417**
Return label (R)	.360	.940	.274	.994
C × R	2.147	.005	1.822	.000
C × G	1.577	.687	1.529	.659
G × R	.181	1.430	.265	1.004
C × G × R	.047	.018	.041	.001
Covariates				
Gender			2.339	.224
Age			.018	1.297
Extraversion			4.157**	1.815
Agreeableness			5.622**	7.516***
Conscientiousness			4.414**	.002
Neuroticism			3.802**	1.148
Openness			.228	.223
Soccer preference			.199	.400
F value	2.230***	1.593	2.568***	1.788**
R-square	.048	.034	.112	.081

* $p < .1$; ** $p < .05$; *** $p < .01$; $N = 320$

model and compared the results between the original research model (without the marker variables) and the common method bias test model (with marker variables). The results indicate that the marker variables had no significant effects on the dependent variables (satisfaction and return intention) or on other effective endogenous variables (utilitarian utility, hedonic utility, and pleasure) (see Online Resource 1.1). In any case, only one relationship between the marker variable and arousal was significant; however, because arousal was non-significant (see Sect. 3.6), this finding does not influence our main conclusions. In addition, the path coefficients between all main contrasts and consumer behavior did not significantly differ between these two models. Therefore, we can conclude that a common method bias did not likely distort the main results of our study.

3.4 Measurement model validation

Our research model contains seven reflective multi-item constructs and six one-item constructs. The quality of the reflective measurement models depends on convergent validity and discriminant validity (Bagozzi and Yi 1988).

To analyze convergent validity, we determined indicator reliability and internal consistency. All the indicator loadings of the reflective multi-item constructs were, at a minimum, significant at the .01 level. For the internal consistency assessment, we examined the composite reliability (CR), Cronbach's alpha, and average variance extracted (AVE) (seeTable 2) (Teo et al. 2003). All the CR indices, as well as the Cronbach's alpha values, met the threshold of .7 (Nunnally et al. 1967). Furthermore, for AVE, all reflective multi-item constructs met Fornell and Larcker's (1981) suggested critical level of .5. In summary, the constructs satisfied all criteria for indicator reliability and internal consistency, in support of convergent validity.

We also analyzed the constructs' discriminant validity by examining whether the square root of the indicators' AVE within any construct was higher than the correlations between it and any other construct (Son and Benbasat 2007). All included constructs met this criterion, thus evidencing discriminant validity (see Online Resource 1.2). Moreover, none of the correlations between any pair of constructs were higher than the threshold value of .9 (Son and Benbasat 2007), and there was no evidence of critically high cross-loadings between the main constructs (see Online Resource 1.3). Therefore, we can conclude that the reflective constructs possessed discriminant validity.

3.5 Results from ANOVA and ANCOVA

We first used ANOVA to test the significant differences in satisfaction and return intention among different package design groups (color, gift, and return label). We then added participants' demographics and personality to the analysis model as covariates (ANCOVA) to test the stability of the results (see Table 4 and Fig. 5).

The results of both analyses showed that an extra gift can significantly influence consumer satisfaction and return intention, while a colorful package only has a significant impact on consumers' return intentions. More specifically, a colorful
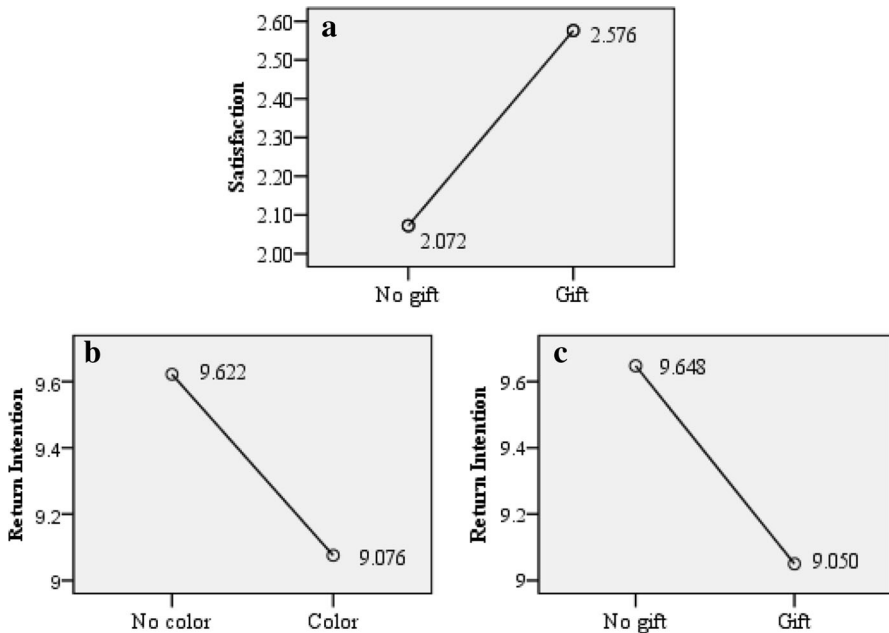
Fig. 5 Influences of the package design on consumers' satisfaction and return intentions

package significantly reduced consumers' return intentions $R(_{cor} = 9.076$ vs. $R_{color} = 9.662$, see Fig. 5b; $F = 3.66$, $p < .1$, see Table 4) compared with a standard brown package, but had no significant impact on consumer satisfaction. Meanwhile, an extra gift in the package increased consumer satisfaction $S(_{gt} = 2.576$ vs. $S_{gift} = 2.072$, see Fig. 5a; $F = 10.685$, $p < .001$, see Table 4) and reduced return intentions $R(_{gt} = 9.050$ vs. $R_{gift} = 9.648$, see Fig. 5c; $F = 4.417$, $p < .05$, see Table 4). These results offer initial evidence for the impact of package design on consumer return behavior.

Our results further showed that a preprinted return label had no significant effect on consumer satisfaction or return intentions ($p > .1$, see Table 4). However, this result might have occurred because European consumers know that their return rights are highly protected by the Consumer Protection Law, and thus the 14-day return policy is already deeply rooted in their decision processes. The other possible reason is that the preprinted return label does not significantly reduce return costs. We also tested the interactions among color, gift, and return label, but none of them were significant.

3.6 Results from PLS analysis

To analyze the package-opening process more thoroughly, we operationalized our model as a structural equation model and estimated it using Smart PLS (v.3.2.1) (Ringle et al. 2015). This method is well suited for exploratory research and shares the modest distributional and sample size requirements of ordinary least squares

linear regression. We also used two models to individually test the cognitive process (without affective reactions) or affective process (without cognitive reactions); the results can be found in the section Online Resource 1.1. To reduce common method bias, we included common control variables for our main dependent variables: age, gender, soccer preference, and personality. The main results appear in Fig. 6.

The squared multiple correlations (R^2) of .39 for satisfaction and .28 for consumers' return intention are high, which means 39% of the variance in satisfaction and 28% of the variance in return intention can be explained by the chosen constructs (Glantz and Slinker 1990). To assess the significance of the path coefficients, we used the bootstrapping procedure implemented in Smart PLS with 1000 resamples. Figure 6 displays the results, with continuous lines representing significant path coefficients and dashed lines indicating non-significant paths.

Package color can positively influence consumer return decisions, as we expected, but surprisingly, our data indicate it only works through the cognitive process via perceived utilities. These results confirm Chebat and Morrin (2007) major finding that, in the realm of consumer behavior, the influence of colors is largely facilitated by cognitive rather than affective mechanisms. Specifically, we found that the perceived utilitarian utility of the blue delivery package is relatively higher than the standard brown package (.166, $p < .01$). In other words, the blue hues associated with a high-value brand can enhance consumers' evaluation of packaged products.

The extra gift significantly increased both the utilitarian utility (.107, $p < .05$) and the hedonic utility (.130, $p < .05$) of the whole package, but showed no significant direct impact on arousal and pleasure. The reason might be that because e-retailers commonly offer extra gifts, consumers may not feel special when
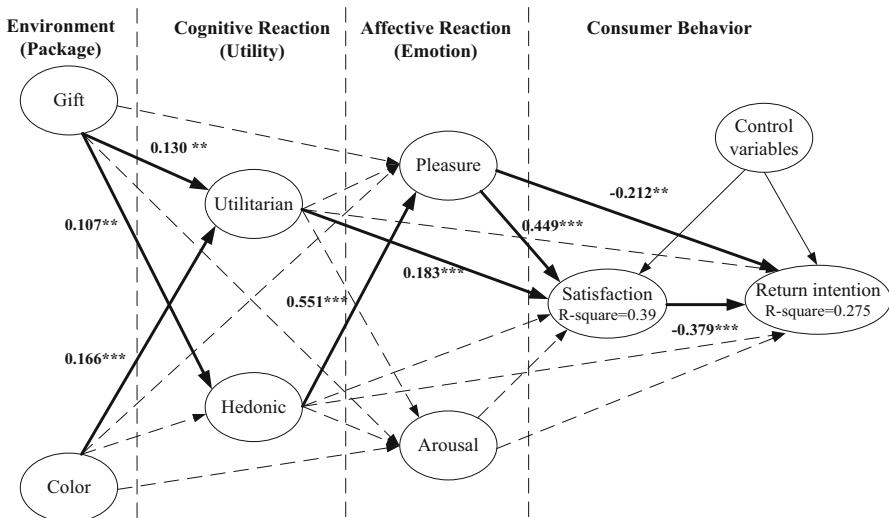
Fig. 6 PLS model and results. *$p < .1$; **$p < .05$; ***$p < .01$; solid arrows indicate significant paths, while dashed arrows indicate insignificant paths

receiving one. At the same time, consumers can easily recognize the utility benefits of extra gifts. When comparing the relative impact of gifts and color, the former works more effectively, but the costs of the latter are significantly lower.

Our results also show that utilitarian and hedonic utility impact the consumers' post-purchase decisions in various ways. Higher utilitarian utility increases consumer satisfaction (.183, $p < .01$), which is consistent with previous empirical findings (e.g., Anderson et al. 2009). In contrast, hedonic utility is positively and strongly related to pleasure (.551, $p < .01$).

In line with our expectations, satisfaction is negatively related to consumer return intention ($-.379$, $p < .01$). In short, the more satisfied consumers are after opening the package, the less return intention they exhibit. The results also indicate that pleasure plays the most crucial role in consumers' post-purchase decisions. Pleasure is the only factor in our research model that can directly increase satisfaction (.449, $p < .01$) and simultaneously decrease return intention ($-.212$, $p < .01$). However, arousal did neither influence satisfaction nor return intention. Indeed, the PLS results revealed that arousal had no significant relationship to any other constructs in our research model.

Furthermore, by using the bootstrapping procedure as a mediation test (Suwelack et al. 2011), we found significant indirect effects of the package design (i.e., extra gifts and colorful packages) on emotions and return intentions (see Table 5), emphasizing the cognitive-affective reactions process. Specifically, we found that extra gifts invoke more pleasure by increasing hedonic utility (.072, $p < .05$, see Table 5). In turn, pleasure can directly and indirectly (via satisfaction, -0.171, $p < .01$, see Table 5) reduce return intentions. In addition, only a colorful package (.030, $p < .1$, see Table 7) can indirectly lead to higher consumer satisfaction, namely by increasing the utilitarian utility. Satisfaction is thus an important mediator, through which utilitarian utility ($-.073$, $p < .05$, see Table 5) and pleasure ($-.171$, $p < .01$, see Table 5) can significantly reduce consumer return intention.

Moreover, we also tested the models that solely included cognitive or affection reactions. The results (see Online Resource 1.1) show that extra gifts and colorful packages can have direct effects on perceived utilities, but not on emotions. The

Table 5 Indirect effects

Indirect effects	Coefficient	SD	T value
Gift → Utilitarian → Satisfaction	.020	.014	1.375
Gift → Hedonic → Pleasure	.072**	.033	2.204
Color → Utilitarian → Satisfaction	.030*	.016	1.869
Utilitarian → Satisfaction → Return intention	$-$ .073**	.032	$-$ 2.194
Hedonic → Pleasure → Return intention	$-$ .111**	.048	$-$ 2.454
Pleasure → Satisfaction → Return intention	$-$ .171***	.047	$-$ 3.534

* $p < .1$; ** $p < .05$; *** $p < .01$

only significant direct effect on emotions is the one from extra gifts on pleasure (.097, $p < .1$), but that might be a result of cognitive reactions like hedonic utility (.072, $p < .01$, see Table 5). Furthermore, we tested the model in reverse order (i.e., affective-cognitive reaction process) and found that our package manipulations did not directly influence affective user reactions (pleasure and arousal). Thus, a cognitive-affective reaction process seems more plausible based on our data.

Among the control variables, only agreeableness had a significantly negative effect on return intentions. In other words, consumers who are kind, sympathetic, cooperative, warm, and considerate are more tolerant of product defects, as might be expected.

3.7 Real return behavior check

Following the experiment's completion, we randomly drew five winners [2 men and 3 women, average age 25.9 years (SD = 10.78)] from the final sample of 320 participants. They received the package as shown in the video of their experimental group. The participants who did not receive a pre-paid DHL label were allowed to email us for a free DHL label (a PDF file). Four of the winners returned the slightly flawed jersey to get a new one and one kept it, which was exactly in line with their stated survey response. This small-number sample may serve as initial evidence that the measured return intention is a reasonable and valid proxy for actual return behavior. This point will be further corroborated in our third study.

4 Study 2: A robustness test with respect to actual return behavior with chocolate bars

To challenge our findings on return intentions, we conducted a robustness test in January 2017 using an experiment with actual return behavior. We invited 394 students from a German university to a controlled, on-campus experiment. This study was designed to estimate the influence of two factors from Study 1 (i.e., extra gifts and colorful packages) on actual return behavior.

First, we presented the participants with a short survey where they rated six chocolate flavors of a well-known German brand on 5-point Likert scales. As a reward, they received a numbered voucher to get the chocolate bar of his/her favorite flavor (e.g., flavor A). Participants had to wait two more hours before they could redeem the voucher (we launched the survey before 11 am and allowed participants to redeem their vouchers after 1 pm). Participants would then receive an envelope containing a chocolate bar but of a non-favorite flavor (e.g., flavor F while flavor A was favored, according to the number on the voucher) in order to create a manipulated product issue. The envelopes were randomly varied by color (brown or blue) and extra gift (with or without a small candy, i.e., mini bag with 5–6 gummy bears). After opening the envelopes, they could find a small note inside saying: "Hi, you can come back for another flavor or receive 1 Euro instead. Have a nice day!"

This study simulates the whole online purchasing process—from ordering (i.e., the choice of their favorite flavor they would get later), a time delay (i.e., 2 h), package receipt (i.e., an envelope), a product issue (i.e., a wrong flavor) and actual return behavior (i.e., change for another flavor or a "refund" in the form of 1 Euro) (see Fig. 7).
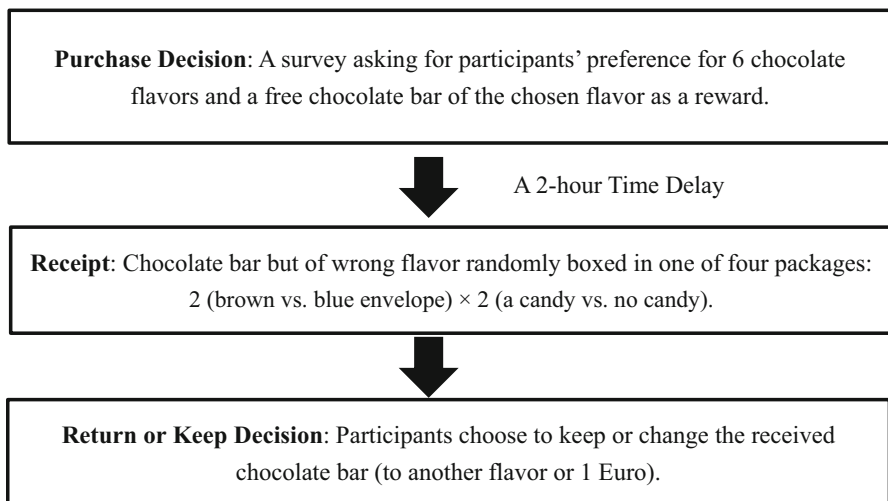
4.1 Method

Of the initial 394 students, 195 (78 females and 117 males) came to redeem their vouchers, while 62 of them returned for a change (see Table 6). Aside from asking for the participants' preference in the survey, we also collected information about participants' gender, the degree of liking chocolate ("Like", 5-point Likert scale with 1 = very little and 5 = very much). We then calculated the standard deviation of the rating scores for six flavors ("Variance") and the difference between the chosen flavor and the received flavor's scores ("Gap"). We coded the package that participants received with "Color" (0 means brown envelope; 1 means blue envelope) and "Gift" (0 means without extra candy; 1 means with extra candy). We coded the participants' actual return behavior as a dummy variable: 0 means the participant kept the "wrong" chocolate bar, while 1 means the participant returned it for another chocolate bar or 1 Euro.

We estimated the following equation with cluster-robust (per day of the experiment) errors: Eq. 1 shows the estimated logistic regression:

$$\text{Prob}(\text{Return Behavior}_i = 1) = F(a'X_i), \qquad (1)$$

where Return Behavior$_i$ is a binary outcome variable with observed values 0 (keep) or 1 (return) which stands for the i_{th} participant's keep or return choice, X_i is a vector collecting related independent variables ("Color$_i$", "Gift$_i$", "Gender$_i$",

---

| **Purchase Decision**: A survey asking for participants' preference for 6 chocolate flavors and a free chocolate bar of the chosen flavor as a reward. |

A 2-hour Time Delay

| **Receipt**: Chocolate bar but of wrong flavor randomly boxed in one of four packages: 2 (brown vs. blue envelope) × 2 (a candy vs. no candy). |

| **Return or Keep Decision**: Participants choose to keep or change the received chocolate bar (to another flavor or 1 Euro). |

Fig. 7 Experimental procedure of Study 2

Table 6 Participants in Study 2 and their return behavior

	Returns	Participants	Return rate (%)
Brown envelope without gift	22	51	43.1
Brown envelope with gift	14	45	31.1
Blue envelope without gift	14	50	28.0
Blue envelope with gift	12	49	24.5
Total	62	195	31.8

"Like$_i$", "Variance$_i$", "Gap$_i$", "Chocolate Chosen$_i$" and "Chocolate Received$_i$") and a potential interaction between "Color" and "Gift" (i.e., "Color$_i$ × Gift$_i$") to X$_i$, and α are the coefficients. $F()$ denotes the logistic distribution. We use Maximum Likelihood (ML) to estimate the model. The odds ratio corresponding to the ith coefficient is calculated by $\psi_i = \exp(\alpha_i)$ and it is approximated with the delta method.

4.2 Results and discussion

Table 7 shows the regression results. These results confirm the positive influences of a colorful package ($\alpha_1 = -.456$, $p < .05$; $\beta_1 = -.789$, $p < .01$) and an extra gift ($\alpha_2 = -.462$, $p < .01$; $\beta_2 = -.819$, $p < .01$) on consumers' actual return behavior. Specifically, when we presented a blue package or an extra candy, the odds of returning decreased by about -37% in both cases. Moreover, there was no interactive effect between a colorful package and an extra gift ($\beta_3 = .737$, $p > .1$), which confirms the ANOVA analysis result of Study 1 (see Table 4).

In sum, the results of Study 2—which includes components of ordering, unpacking and actual return behavior—fully support the initial findings from Study 1. Thus, we can claim that our conclusions are not just based on artificial effects, but rather demonstrate good robustness and high validity in other experimental settings.

5 Study 3: An analysis of transactional data on Taobao.com

As a follow-up study, we analyzed transactional field data to examine whether package design (i.e., color, gift, and coupon) is related to e-retailers' return rates. We analyzed transactional data from the largest online shopping platform in China, Taobao.com, and used web shops' return rates as the dependent variable.

5.1 Method

Taobao.com was founded in 2003 and currently belongs to the Alibaba group. The site had 500 million registered users by the end of 2013 (Ye et al. 2013). We drew a random sample of 400 e-retailers from a large sample pool (2826 e-retailers from

Table 7 Regression results of Study 2

Return behavior (0/1)	Measurement	Without interaction			With interaction		
		Coef.	Odds ratio	SE of Coef.	Coef.	Odds ratio	SE of Coef.
Intercept	0/1	.677	1.969	.785	.892	2.439	.797
Color	0/1	− .456	.634	.192**	− .789	.454	.209***
Gift	0/1	− .462	.630	.067***	− .819	.441	.303***
Color* gift	0/1				.737	2.089	.476
Gender	0/1	− .082	.921	.342	− .034	.967	.362
Like	Degree of liking chocolate (scale 1–5)	− .179	.836	.134	− .216	0.806	.133
Variance	SD of six Flavors' scores	.123	1.130	.775	.167	1.181	.729
Gap	Score$_{choose}$–Score$_{received}$ (range 0–4)	.103	1.108	.087	.104	1.109	.094
Chocolate chosen	Categorical variable						
Flavor A		Fixed	1		Fixed	1	
Flavor B		− .285	.752	.111**	− .305	.737	.104***
Flavor C		− .974	.378	.258***	− 1.044	.352	.266***
Flavor D		− .717	.488	.629	− .778	.459	.676
Flavor E		− .785	.456	.794	− .743	.476	.785
Flavor F		− 1.612	.200	1.187	− 1.511	.221	1.208
Chocolate Offered	Categorical variable						
Flavor A		Fixed	1		Fixed	1	
Flavor B		− 2.190	.112	.438***	− 2.267	.104	.480***
Flavor C		− .705	.494	.723	− .740	.477	.732
Flavor D		− 1.239	.290	.929	− 1.309	0.270	1.002
Flavor E		.243	1.275	.555	.265	1.303	.536
Flavor F		− .149	.862	.345	− .202	.817	.305
Pseudo R-square		.117			.122		

* $p < .1$; ** $p < .05$; *** $p < .01$; $N = 195$

Taobao.com) in August 2014 from the categories "digital products" and "clothing". We used these industries because of their relatively high return rates and the substantial variation in their e-retailers' delivery package designs. With the links stored in our database, we also collected information about these e-retailers' delivery packages.

We obtained information about each e-retailer's delivery package, including the color, extra gifts, and coupons, by using numerous consumer comments (both text and photo) and retailer product demonstrations. At Taobao.com, it is normal for consumers to upload photo reviews of their received products, including both the

outside appearance and its contents. Meanwhile, fierce competition on the platform encourages e-retailers to describe as many product details as they can, including delivery packages, coupons and gifts. Thus, we were able to obtain detailed package information from many e-retailers' websites. We restricted our sample to e-retailers where complete package information was obtainable. We coded "Gift" as 0 if there was no gift included and "Gift" as 1 if the particular e-retailer provided at least one gift. We coded "Color" as 0 if the package box was the typical light brown and 1 if the package box was a different color (e.g., red, blue, pink). We excluded e-retailers that used plastic bags to ship orders. We also coded "Coupon" as 0 if there was no coupon and 1 if there was at least one coupon in the delivery package. Two Chinese students coded the data, which resulted in a sample of 108 shops for which we could reliably reconstruct the delivery package design (see Table 8). We should note that e-retailers do not provide preprinted return labels in China, so we were not able to examine the effect of return costs in this study.

Our dependent variable is the e-retailer's return rate, which is calculated by the platform using the number of completed returns (calculated by the platform's internal return system) divided by the number of deals (calculated by the platform's trading system) in the observation month. This value is automatically generated by Taobao.com and is displayed on every e-retailers' public rating page. We also considered other related factors as control variables: The leniency of the return policy is measured by "Guarantee Money", which platform operators collect in advance in order to enforce e-retailers' compliance with their promised return policy and offer refunds in case of disputes between e-retailers and consumers. "Review score" is measured by a customer-generated score (ranging from 1 = "very bad" to 5 = "very good") that rates a shop's service quality in the recent half year. The "Industry Return Rate" measures the average return rate across a given industry. The variable "Market" represents the two markets served on Taobao.com: a business-to-consumer market coded as 1 and a consumer-to-consumer market coded as 0. Based on the range of return rates (0–1), we estimated a regression model for fractional response variables with logit distribution (called "fractional logistic model" by STATA) with robust standard errors to avoid a strong influence of outliers. We thus estimated the following equation:

Table 8 Descriptive statistics of the valid sample in Study 3

Variable (unit)	Return rate (%)	Guarantee money (Chinese Yuan)	Service score (range 1–5)
Mean	6.33	65,400	4.77
SD	9.29	38,663	.083

	Gift	Coupon	Colorful package
Number of shops with gift/coupon/colorful package	64	15	28
Number of shops without gift/coupon/colorful package	44	93	80

$N = 108$

$$\text{Prob}(\text{Return Rate}_i) = F('Z_i), \tag{2}$$

where Return Rate$_i$ stands for the i-th e-retailer's return rate (an aggregate percentage figure, i.e. between 0 and 100%), Z_i represents a vector for all related independent variables ("Gift", "Coupon", "Color", "Gurantee Money", "Review Score", "Industry Return Rate" and "Market"), and β' is the corresponding vector of the coefficients. We use the maximum quasi-likelihood[3] method to estimate the regression model with fractional dependent variable. The odds ratio corresponding to the ith coefficient is $\psi_i = \exp(\beta_i)$ which is approximated with the delta method.

6 Results and discussion

Table 9 shows the regression results, which reveal that extra gifts and colorful packages have significant and negative impacts on shops' return rates ($\gamma_1 = -.721$, $p < .05$; $\gamma_3 = -.843$, $p < .01$). In other words, a well-designed delivery package can help reduce return intentions. However, our estimation results indicate that coupons do not significantly reduce return rates ($\gamma_2 = .049$, n.s.). We surmise that coupons may only be related to consumers' repurchase decisions and not to their return decisions. Another possible explanation is that the prevalence of e-coupons on the Internet makes coupons less useful. In addition, the data show that improving service and product quality ($\gamma_5 = -2.417$, $p < .1$) is the most effective way for e-retailers to decrease return rates.

7 Discussion and implications

Previous research has acknowledged that various factors (e.g., product quality, product demonstration, return policies), working through a cognitive reaction process, can significantly impact consumer return behavior during the two-stage decision process of online purchasing (Anderson et al. 2009; Frischmann et al. 2012; Petersen and Kumar 2009; Wood 2001). However, no study has yet examined how consumer return decisions are influenced by a delivery package's appearance (e.g., the color) and contents (e.g., extra gifts, preprinted return labels), or the affective action process. To close these gaps and identify useful packaging strategies, we combined the synergistic properties of two controlled experiments and an observational field study, which together offer robust insights into how the delivery package affects return intentions.

Our studies offer three key findings. The first is that the appearance (i.e., colorful package) and composition (i.e., extra gift) of the delivery package can positively influence consumers' return decisions, at least when said package is a part of the integral purchasing experience. This conclusion was initially supported by the

---

[3] The quasi-likelihood function was introduced by Robert Wedderburn in 1974 to describe a function which has similar properties to the log-likelihood function, except that a quasi-likelihood function is not the log-likelihood corresponding to any actual probability distribution. Quasi-likelihood models can be fitted using a straightforward extension of the algorithms used to fit generalized linear models.

Table 9 Results of regression model for fractional response variables

Return rate	Coef.	Odds ratio	SE of Coef.	$p > |z|$
Intercept	7.759	2341.707	6.217	0.212
Gift	− .721**	0.486	0.317	0.023
Coupon	.049	1.050	0.327	0.882
Color	− .843***	0.430	0.242	0.000
Guarantee money	− .00001	1.000	0.000	0.167
Review score	− 2.417*	0.089	1.340	0.071
Industry return rate	.282***	1.326	0.058	0.000
Market	.573	1.775	0.554	0.301
Pseudo R-square	0.0765			

$N = 108$. * $p < .1$; ** $p < .05$; *** $p < .01$

results of a controlled experiment (Study 1), then confirmed by the results of an experiment involving real return behavior (Study 2), and finally supplemented by our analysis of a cross-sectional, transactional data set (Study 3). Second, our results reveal a two-stage reaction process ("high-road" cognitive–affective) when consumers open the delivery package. These cognitive–affective reactions can also more thoroughly explain how the delivery package influences consumers' return behavior. Third and finally, we show that perceived pleasure is the only reaction that can directly influence both consumer satisfaction and return intentions.

7.1 Theoretical contributions

To the best of our knowledge, this paper is the first to examine the entire chain of the two-stage, "high-road" cognitive–affective reaction process in the context of consumer return behavior, which contributes an integrated and holistic perspective to the return behavior research field. In prior studies, researchers have mainly focused on the cognitive process at the neglect of the affective process. In this study, we confirm that perceived utility has an impact on return intention, but more in an indirect way, i.e. via satisfaction (i.e., utilitarian utility). However, the "high-road" affective reaction (i.e., pleasure perceptions), rather than the cognitive reaction, is directly responsible for consumers' return decisions. Moreover, the hedonic utility which with utilitarian utility constitutes an integral part of the perceived utility (Chandon et al. 2000) has an indirect impact on return intention only via pleasure. Thus, we recommend that future research look beyond consumers' cognitive reactions (e.g., perceived utilitarian utility and expectation gap) and pay closer attention to hedonic utility and the influence of consumers' affective reactions (i.e., emotions).

In addition, this article is one of the first to highlight the importance of the package-opening process. Previous research has explored how a consumer's return decision is influenced by the purchase stage (Petersen and Kumar 2009), a delay

period (Bechwati and Siegal 2005), and the final user experience, but has largely ignored the short but critical moment of unpacking. After purchasing, consumers may have a vague impression of their purchase decision, but they are still easily affected when making their return decisions. A well-designed delivery package makes a first and important impression and can ultimately influence consumers' return decisions. As a result, we believe that models about consumers' post-purchase decision-making should factor in delivery package design.

Furthermore, on the one hand, this work adds color and gift into the long list of influential factors to consumer return behavior. We find that both, color and extra gifts, can significantly decrease consumer's return intentions by adding extra perceived utilities. On the other hand, this study also enriches color and gift research by demonstrating their positive influences on the post-purchase behavior. By integrating literatures on color psychology, promotion, and consumer return behavior, we discovered that colors and extra gifts can influence both purchase and return decisions.

7.2 Practical contributions

Given our key findings, we suggest that e-retailers focus on the delivery package design to lower consumers' return intentions. Choosing a suitable color for delivery packages and offering extra gifts can potentially mitigate return rate problems. With respect to cost efficiency, colorful packages seem more promising because they are cheaper to implement than extra gifts. Specifically, the price of a colorful packaging carton is around 0.136 Euro (1 Yuan, according to Alibaba.com). There is no difference between normal (brown) cartons and colorful ones (even the designed one with words and graphs) when the order quantity exceeds 1000. As a result, when e-retailers can order more than 1000 cartons at one time, even a small gift (like the candy used in Study 2) is more expensive than a designed package. In addition, color can help increase brand familiarity (Labrecque and Milne 2012). For these reasons, we recommend using colorful packages.

Moreover, e-retailers should pay more attention to consumers' emotions (e.g., pleasure perceptions). We used colorful delivery packages and extra gifts to generate positive emotions, but other add-ins (e.g., a note with jokes or wishes) might also be useful. On the flip side, e-retailers might mitigate consumers' negative emotions by offering better post-purchase service.

Granted, the positive effects of package design on emotions (i.e., positive surprise) might wear off over time or with widespread use, but the positive effects on perceived utility are sustainable. Thus, we highly encourage e-retailers to think about a suitable package design strategy.

7.3 Limitations

This study comes with several limitations. First, although we carefully chose the products in our experiment—a jersey of the German national football team (in Study 1) and a chocolate bar (in Study 2)—we cannot easily generalize their effects to all other products. However, because we compared relative differences across the

experimental groups, this limitation should not constitute a severe problem for our analyses.

Second, the controlled experiments only included one hue for the colorful package (i.e., blue in Study 1 and 2) and two gifts (Nivea cream in Study 1 and candy in Study 2), which limits their generalizability. For example, the PLS results showed no significant relationship between a blue package and consumers' perceived hedonic utility, but the result might change for a yellow package. Moreover, even the same hue (i.e., blue) might have different effects in different cultures. Future research could explore these possibilities in greater deail. Still, our study provides significant evidence that a well-designed delivery package can positively influence consumer return intention. E-retailers should experiment with different colors and/or gifts in order to achieve a suitable and memorable delivery package.

Third, the between-subjects experiment design in Studies 1 and 2 led to a limited sample size for each experimental group. In both experimental studies, the number of participants in each group did not exceed 50, which may have influenced the accuracy of our study. Furthermore, the general limitations of a cross-sectional study might have biased the results of our field study (Study 3). Some unobserved factors (such as the e-retailer's brand strength) might influence e-retailers' return rate, but could not be included in our regression models. In fairness, the consistent results of three studies should imbue our conclusions with some confidence. Nevertheless, future studies should further explore our research questions by using panel data analysis. Because panel-econometric approaches can better control for confounding effects, they may be able to corroborate our findings and even deduce the influence of changing one's package strategy.

In sum, this article offers a new research perspective on consumer return behavior, one that found consistent results across Chinese and German data. However, there is a need for further research into how consumer return behavior is influenced by the product category or different color hues and gifts. Scholars could also explore the interaction between different customer segments and package design, such as how different color hues present variable effects across cultures.

References

Anderson, Eric T, Karsten Hansen, DI Simester, and Lei K Wang. 2009a. How price affects returns: The perceived value and incremental customer effects. *Unpublished Manuscript, Northwestern University, Evanston, IL.*

Anderson, Eric T., Karsten Hansen, and Duncan Simester. 2009b. The option value of returns: theory and empirical evidence. *Marketing Science* 28 (3): 405–423.

Aurier, Philippe, and Guergana Guintcheva. 2014. Using affect–expectations theory to explain the direction of the impacts of experiential emotions on satisfaction. *Psychology & Marketing* 31 (10): 900–913.

Bagozzi, Richard P., and Youjae Yi. 1988. On the evaluation of structural equation models. *Journal of the Academy of Marketing Science* 16 (1): 74–94.

Banjo, Shelly. 2013. Rampant returns plague e-retailers. *The Wall Street Journal.* http://online.wsj.com/news/articles/SB10001424052702304773104579270260683155216. Accessed 15th May 2015.

Bechwati, Nada Nasr, and Wendy Schneier Siegal. 2005. The Impact of the Prechoice Process on Product Returns. *Journal of Marketing Research* 42 (3): 358–367.

Berkowitz, Leonard. 1993. *Aggression: Its causes, Consequences, and Control.* Mcgraw-Hill Book Company.

Bonifield, Carolyn, and Catherine Cole. 2007. Affective responses to service failure: anger, regret, and retaliatory versus conciliatory responses. *Marketing Letters* 18 (1–2): 85–99.

Chandon, Pierre, Brian Wansink, and Gilles Laurent. 2000. A benefit congruency framework of sales promotion effectiveness. *Journal of Marketing* 64 (4): 65–81.

Chang, En-Chung, Yilin Lv, Ting-Jui Chou, Qingwen He, and Zhuozhao Song. 2014. Now or later: delay's effects on post-consumption emotions and consumer loyalty. *Journal of Business Research* 67 (7): 1368–1375.

Chebat, Jean-Charles, and Maureen Morrin. 2007. Colors and cultures: exploring the effects of mall décor on consumer perceptions. *Journal of Business Research* 60 (3): 189–196.

Chitturi, Ravindra, Rajagopal Raghunathan, and Vijay Mahajan. 2007. Form versus function: how the intensities of specific emotions evoked in functional versus hedonic trade-offs mediate product preferences. *Journal of Marketing Research* 44 (4): 702–714.

Choi, Sunmee, and Anna S. Mattila. 2008. Perceived controllability and service expectations: influences on customer reactions following service failure. *Journal of Business Research* 61 (1): 24–30.

Crowley, Ayn E. 1993. The two-dimensional impact of color on shopping. *Marketing Letters* 4 (1): 59–69.

DeWitt, Tom, Doan T. Nguyen, and Roger Marshall. 2008. Exploring customer loyalty following service recovery the mediating effects of trust and emotions. *Journal of Service Research* 10 (3): 269–281.

Elliot, Andrew J., and Markus A. Maier. 2007. Color and psychological functioning. *Current Directions in Psychological Science* 16 (5): 250–254.

Eurostat. 2014. Anteil der Online-Käufer in folgenden Altersgruppen in Europa im Jahr 2014. http://de.statista.com/statistik/daten/studie/76300/umfrage/altersverteilung-der-online-kaeufer-in-deutschland/. Accessed 30th Feb 2016.

Finn, Adam. 2005. Reassessing the foundations of customer delight. *Journal of Service Research* 8 (2): 103–116.

Frischmann, Tanja, Oliver Hinz, and Bernd Skiera. 2012. Retailers' use of shipping cost strategies: free shipping or partitioned prices? *International Journal of Electronic Commerce* 16 (3): 65–88.

Garretson, Judith A., and Scot Burton. 2005. The role of spokescharacters as advertisement and package cues in integrated marketing communications. *Journal of Marketing* 69 (4): 118–132.

Gelbrich, Katja. 2010. Anger, frustration, and helplessness after service failure: coping strategies and effective informational support. *Journal of the Academy of Marketing Science* 38 (5): 567–585.

Glantz, Stanton A., and Bryan K. Slinker. 1990. *Primer of applied regression and analysis of variance.* New York: McGraw-Hill Book Company.

Heilman, Carrie M., Kent Nakamoto, and Ambar G. Rao. 2002. Pleasant surprises: consumer response to unexpected in-store coupons. *Journal of Marketing Research* 39 (2): 242–252.

Heiman, Amir, Bruce McWilliams, and David Zilberman. 2001. Demonstrations and money-back guarantees: market mechanisms to reduce uncertainty. *Journal of Business Research* 54 (1): 71–84.

Hibbert, Sally A., Maria G. Piacentini, and Margaret K. Hogg. 2012. Service recovery following dysfunctional consumer participation. *Journal of Consumer Behaviour* 11 (4): 329–338.

Institut für Demoskopie Allensbach 2001. Welche ist Ihre Lieblingsfarbe? http://de.statista.com/statistik/daten/studie/50/umfrage/eigene-lieblingsfarbe/. Accessed 30th Oct 2014.

Institut für Demoskopie Köln 2015. How high is the average rate of return of goods in your shop per year? (Germany, 2014). http://www.statista.com/statistics/327314/share-of-returns-in-online-shopping-regarding-fashion-and-consumer-electronics/. Accessed 16th Aug. 2015.

Jacobs, Laurence, Charles Keown, Reginald Worthley, and Kyung-Il Ghymn. 1991. Cross-cultural colour comparisons: global marketers beware! *International Marketing Review* 8 (3): 21–33.

Janakiraman, Narayan, and Lisa Ordóñez. 2012. Effect of effort and deadlines on consumer product returns. *Journal of Consumer Psychology* 22 (2): 260–271.

Labrecque, Lauren I., and George R. Milne. 2012. Exciting red and competent blue: the importance of color in marketing. *Journal of the Academy of Marketing Science* 40 (5): 711–727.

Labrecque, Lauren I., Vanessa M. Patrick, and George R. Milne. 2013. The marketers' prismatic palette: a review of color research and future directions. *Psychology & Marketing* 30 (2): 187–202.

Ladhari, Riadh. 2007. The effect of consumption emotions on satisfaction and word-of-mouth communications. *Psychology & Marketing* 24 (12): 1085–1108.

LeDoux, Joseph E. 1995. Emotion: clues from the brain. *Annual Review of Psychology* 46: 209–235.

Liu, Hsin-Hsien, and Hsuan-Yi Chou. 2015. The effects of promotional frames of sales packages on perceived price increases and repurchase intentions. *International Journal of Research in Marketing* 32 (1): 23–33.

Madden, Thomas J., Kelly Hewett, and Martin S. Roth. 2000. Managing images in different cultures: a cross-national study of color meanings and preferences. *Journal of International Marketing* 8 (4): 90–107.

Mazaheri, Ebrahim, Marie-Odile Richard, and Michel Laroche. 2012. The role of emotions in online consumer behavior: a comparison of search, experience, and credence services. *Journal of Services Marketing* 26 (7): 535–550.

Mazaheri, Ebrahim, Marie Odile Richard, Michel Laroche, and Linda C. Ueltschy. 2014. The influence of culture, emotions, intangibility, and atmospheric cues on online behavior. *Journal of Business Research* 67 (3): 253–259.

McClelland, James L. 1988. Connectionist models and psychological evidence. *Journal of Memory and Language* 27 (2): 107–123.

Mehrabian, Albert, and James A. Russell. 1974. *An approach to environmental psychology*. Cambridge: MIT Press.

Mehta, Ravi, and Rui Juliet Zhu. 2009. Blue or red? Exploring the effect of color on cognitive task performances. *Science* 323 (5918): 1226–1229.

Meyers-Levy, Joan, and Laura A. Peracchio. 1995. Understanding the effects of color: how the correspondence between available and required resources affects attitudes. *Journal of Consumer Research* 22 (3): 121–138.

Nunnally, Jum C., Ira H. Bernstein, and Jos M.F. ten Berge. 1967. *Psychometric theory*. New York: McGraw-Hill Book Company.

Oliver, Richard L. 1993. Cognitive, affective, and attribute bases of the satisfaction response. *Journal of Consumer Research* 20 (3): 418–430.

Petersen, J.A., and V. Kumar. 2009. Are product returns a necessary evil? Antecedents and consequences. *J Market* 73 (3): 35–51.

Podsakoff, Philip M., Scott B. MacKenzie, Jeong-Yeon Lee, and Nathan P. Podsakoff. 2003. Common method biases in behavioral research: a critical review of the literature and recommended remedies. *Journal of Applied Psychology* 88 (5): 879–903.

Puccinelli, Nancy M., Rajesh Chandrashekaran, Dhruv Grewal, and Rajneesh Suri. 2013. Are men seduced by red? The effect of red versus black prices on price perceptions. *Journal of Retailing* 89 (2): 115–125.

Rammstedt, Beatrice, and Oliver P. John. 2007. Measuring personality in one minute or less: a 10-item short version of the Big Five Inventory in English and German. *Journal of Research in Personality* 41 (1): 203–212.

Reichheld, Frederick F. 2003. The one number you need to grow. *Harvard Business Review* 81 (12): 46–55.

Ringle, Christian M, Sven Wende, and Jan-Michael Becker. 2015. SmartPLS 3. *Boenningstedt: SmartPLS GmbH*. http://www.smartpls.com.

Roggeveen, Anne L., Ronald C. Goodstein, and Dhruv Grewal. 2014. Improving the effect of guarantees: the role of a retailer's reputation. *Journal of Retailing* 90 (1): 27–39.

Rönkkö, Mikko, and Jukka Ylitalo. 2011. PLS marker variable approach to diagnosing and controlling for method variance. Paper presented at the Proceedings of the International Conference on Information Systems Shanghai.

Sahoo, Nachiketa, Chrysanthos Dellarocas, and Shuba Srinivasan. 2015. The impact of online product reviews on product returns. In *Boston U. School of Management Research Paper No. 2491276*

Samson, Alain. 2006. Understanding the buzz that matters: negative vs positive word of mouth. *International Journal of Market Research* 48 (6): 647–657.

Scherer, Klaus R, Angela Schorr, and Tom Johnstone. 2001. *Appraisal processes in emotion: Theory, methods, research*. Oxford University Press.

Schoefer, Klaus. 2010. Cultural moderation in the formation of recovery satisfaction judgments: a cognitive-affective perspective. *Journal of Service Research* 13 (1): 52–66.

Shiv, Baba, and Alexander Fedorikhin. 1999. Heart and mind in conflict: the interplay of affect and cognition in consumer decision making. *Journal of Consumer Research* 26 (3): 278–292.

Smith, Amy K., and Ruth N. Bolton. 2002. The effect of customers' emotional responses to service failures on their recovery effort evaluations and satisfaction judgments. *Journal of the Academy of Marketing Science* 30 (1): 5–23.

Son, Jai-Yeol, and Izak Benbasat. 2007. Organizational buyers' adoption and use of B2B electronic marketplaces: efficiency-and legitimacy-oriented perspectives. *Journal of Management Information Systems* 24 (1): 55–99.

Suwelack, Thomas, Jens Hogreve, and Wayne D. Hoyer. 2011. Understanding money-back guarantees: cognitive, affective, and behavioral outcomes. *Journal of Retailing* 87 (4): 462–478.

Teo, Hock-Hai, Kwok Kee Wei, and Izak Benbasat. 2003. Predicting intention to adopt interorganizational linkages: an institutional perspective. *MIS Quarterly* 27 (1): 19–49.

Voss, Kevin E., Eric R. Spangenberg, and Bianca Grohmann. 2003. Measuring the hedonic and utilitarian dimensions of consumer attitude. *Journal of Marketing Research* 40 (3): 310–320.

Wood, Stacy L. 2001. Remote Purchase Environments: the influence of return policy leniency on two-stage decision processed. *Journal of Marketing Research* 32 (2): 157–169.

Ye, Qiang, Xu Min, Melody Kiang, Wu Weifang, and Fangfang Sun. 2013. In-depth analysis of the seller reputation and price premium relationship: a comparison between eBay US and Taobao China. *Journal of Electronic Commerce Research* 14 (1): 1–10.

Zeltner, Philip M. 1975. *John Dewey's aesthetic philosophy*. Amsterdam: John Benjamins Publishing Company.

Zhou, Wenyan, and Oliver Hinz. 2016. Determining profit-optimizing return policies—a two-step approach on data from Taobao.com. *Electronic Markets* 26(2):103-114.

Publisher's Note