

# **The Moment of Delivery in Online Retailing: How Courier Types and Inoculation Messages Affect Reactions to the Last-Mile Experience**

Christopher Funke <sup>a</sup>, Gianfranco Walsh <sup>a</sup>, Tobias Schaefers <sup>b, c</sup>

<sup>a</sup> Institute of Marketing and Management, Leibniz University of Hannover, Hannover, Germany

<sup>b</sup> Bielefeld School of Business, HSBI – Bielefeld University of Applied Sciences and Arts, Bielefeld, Germany

<sup>c</sup> Department of Marketing, Copenhagen Business School, Frederiksberg, Denmark

<https://doi.org/10.1177/10946705261464334>

## **WEB APPENDIX**

|                                                                           |           |
|---------------------------------------------------------------------------|-----------|
| <b>WEB APPENDIX A: PRE-STUDY .....</b>                                    | <b>2</b>  |
| <b>WEB APPENDIX B: STIMULUS MATERIAL (Studies 1a–1c) .....</b>            | <b>10</b> |
| <b>WEB APPENDIX C: STIMULUS MATERIAL (Study 1a, pretest).....</b>         | <b>11</b> |
| <b>WEB APPENDIX D: ITEMS AND RELIABILITY MEASURES (Studies 1–2) .....</b> | <b>12</b> |
| <b>WEB APPENDIX E: MODERATED MEDIATION ANALYSIS (Study 1a).....</b>       | <b>13</b> |
| <b>WEB APPENDIX F: STIMULUS MATERIAL (Study 2).....</b>                   | <b>16</b> |
| <b>REFERENCES.....</b>                                                    | <b>17</b> |

## WEB APPENDIX A: PRE-STUDY

### Sample

To assess our baseline assumption that the moment of delivery (MoD) affects how customers evaluate online retailers and how they share their experience with others, we analyzed publicly available customer reviews. We randomly selected 28 online retailers, representing 13 different industries (e.g., clothing, consumer electronics, food, online pharmacy, books, bicycles) listed on Trusted Shops (<https://www.trustedshops.de>), a German independent platform that online retailers use to collect customer feedback and reviews. We chose this platform because it ensures that only customers who made a purchase can provide a review and because it represents recent service experiences, as only reviews published within the past 12 months are available.

For the reviews of each online retailer, we automatically captured the star rating, ranging from one to five stars. Additionally, if customers provided a written review<sup>1</sup>, we extracted the full text. The compiled data consist of 59,179 customer reviews with an average star rating of 4.751 (see Table WA1), ranging from 4.165 to 4.934 across the different retailers (see Table WA2). Of all the reviews, 35,822 (60.5%) include a written review.

**Table WA1. Star Rating Distribution and Descriptives**

| Star rating | 1               | 2            | 3             | 4               | 5                 | Total            | Mean<br>(SD)    |
|-------------|-----------------|--------------|---------------|-----------------|-------------------|------------------|-----------------|
| Full sample | 2.5%<br>(1,503) | .8%<br>(487) | 1.6%<br>(942) | 9.1%<br>(5,368) | 86.0%<br>(50,879) | 100%<br>(50,879) | 4.751<br>(.757) |

**Table WA2. Star Rating Descriptives by Online Retailer**

| Retailer | 1      | 2       | 3      | 4      | 5      | 6      | 7       | 8       | 9      | 10     | 11     | 12     | 13     | 14     |
|----------|--------|---------|--------|--------|--------|--------|---------|---------|--------|--------|--------|--------|--------|--------|
| M        | 4.582  | 4.810   | 4.872  | 4.558  | 4.869  | 4.754  | 4.892   | 4.165   | 4.853  | 4.857  | 4.824  | 4.797  | 4.683  | 4.720  |
| (SD)     | (.814) | (.665)  | (.494) | (.957) | (.529) | (.734) | (.402)  | (1.482) | (.571) | (.537) | (.616) | (.642) | (.715) | (.744) |
| n        | 2,700  | 2,620   | 2,740  | 405    | 2,680  | 183    | 2,840   | 2,175   | 900    | 2,680  | 2,800  | 1,569  | 2,388  | 1,149  |
| Retailer | 15     | 16      | 17     | 18     | 19     | 20     | 21      | 22      | 23     | 24     | 25     | 26     | 27     | 28     |
| M        | 4.897  | 4.370   | 4.707  | 4.802  | 4.735  | 4.770  | 4.535   | 4.718   | 4.851  | 4.848  | 4.713  | 4.934  | 4.847  | 4.887  |
| (SD)     | (.433) | (1.293) | (.807) | (.705) | (.803) | (.680) | (1.127) | (.800)  | (.597) | (.513) | (.816) | (.309) | (.487) | (.409) |
| n        | 213    | 1,594   | 2,520  | 1,620  | 2,720  | 2,780  | 2,560   | 2,560   | 2,800  | 1,123  | 2,840  | 2,760  | 2,660  | 2,600  |

<sup>1</sup> Trusted Shops ([www.trustedshops.de](https://www.trustedshops.de)) does not provide any default structure for a written review.

## Method

We content analyzed all 35,822 review texts, using two analytical approaches, namely dictionaries and large language models (Villarroel Ordenes et al. 2025).

*Dictionary-based text categorization.* For the first approach, we used a customized dictionary of delivery-related terms to identify whether or not a review text refers to the delivery. For the dictionary, we compiled a list of eight terms that resemble a reference to the delivery, as well as eight terms that capture references to specific logistics service providers (see Table WA3). We then automatically categorized every review as either containing or not containing each of the dictionary terms.

**Table WA3. Dictionary Terms**

| Original term | Examples                                                                   | Translation/meaning                                                     |
|---------------|----------------------------------------------------------------------------|-------------------------------------------------------------------------|
| “liefer”      | <i>Lieferung</i><br><i>liefern</i><br><i>geliefert</i><br><i>lieferbar</i> | delivery<br>to deliver<br>delivered<br>available, deliverable, in stock |
| “versand”     | <i>Versand</i><br><i>versandt</i>                                          | shipment<br>shipped, dispatched, sent                                   |
| “versend”     | <i>versendet</i><br><i>versenden</i>                                       | shipped, dispatched, sent<br>to ship, to dispatch, to send              |
| “zustell”     | <i>Zustellung</i><br><i>zustellen</i>                                      | delivery<br>to deliver                                                  |
| “transport”   | <i>Transport</i><br><i>transportieren</i><br><i>transportiert</i>          | transport, shipment<br>to transport<br>transported                      |
| “send”        | <i>Sendung</i><br><i>senden</i><br><i>gesendet</i>                         | Shipment, parcel, mailing<br>to send<br>sent, dispatched, mailed        |
| “bote”        | <i>Paketbote</i><br><i>Postbote</i>                                        | parcel carrier<br>postman, mailman                                      |
| “dhl”         | <i>DHL</i>                                                                 | Logistics service provider                                              |
| “gls”         | <i>GLS</i>                                                                 | Logistics service provider                                              |
| “ups”         | <i>UPS</i>                                                                 | Logistics service provider                                              |
| “dpd”         | <i>DPD</i>                                                                 | Logistics service provider                                              |
| “hermes”      | <i>Hermes</i>                                                              | Logistics service provider                                              |
| “fedex”       | <i>FedEx</i>                                                               | Logistics service provider                                              |
| “dachser”     | <i>Dachser</i>                                                             | Logistics service provider                                              |

|          |                                         |                             |
|----------|-----------------------------------------|-----------------------------|
| “post”   | <i>Deutsche Post</i><br><i>CityPost</i> | Logistics service providers |
| “spedit” | <i>Spedition</i><br><i>Spediteur</i>    | shipping company<br>carrier |

*Note:* German terms were used as all the extracted reviews were written in German.

*Large Language Model-based text categorization.* For the second approach, we relied on recent advances in generative artificial intelligence (AI) and large language models (LLMs) to automatically content analyze each review text. Our approach is similar to the LLM-assisted inductive categorization (LAIC) developed by Arlinghaus et al. (2024). In contrast to LAIC, however, we did not follow an inductive path but employed a deductive identification and evaluation of specific contents. Using an application programming interface provided by OpenAI (<https://openai.com/api/>), we instructed its gpt-4.1-2025-04-14 model to content analyze each review, with the following instructions:

*“You are a text analyst who analyzes customer reviews of online stores. Analyze each of the review texts and answer the following questions: 1. Is delivery or shipping mentioned? (Answer: Yes or No) 2. If so, is delivery or shipping rated positively, negatively, mixed, or neutral? (Answer: positive, negative, mixed, neutral).”*

The procedure thus created two variables, a binary one capturing whether a review text contains any reference to the delivery process, and a categorical variable capturing the way in which the delivery experience is described. Such a so-called zero-shot classification “eliminates the need for explicit model training and makes LLMs a versatile method for automated text analysis such as aspect-based sentiment analysis” (Villarroel Ordenes et al. 2025, p. 385).

Both approaches provided us with different categorization variables for each review, which we leverage in comparative analyses of the star ratings. At the same time, it is important to consider the differences of the two approaches (Villarroel Ordenes et al. 2025).

On the one hand, the dictionary-based analysis represents a relatively simple and easily replicable approach. On the other hand, the LLM-based categorization can account for more contextual cues, reduces the risk of false negative classification due to missing dictionary terms, and exhibits “greater accuracy in measuring emotional states than dictionary methods (Villarroel Ordenes et al. 2025, p. 386).

*Statistical analyses.* With regards to the analyses of the categorized reviews, three aspects need to be considered. First, the identified categories differ substantially in their group sizes and variances, as is evident in Tables WA4 and WA5. This likely reduces the robustness of parametric analytical methods, such as t-tests or analysis of variance. Therefore, we employed non-parametric tests for direct comparisons.

Second, the star rating data is heavily skewed, as is evident in Table WA1 (e.g., 86% of all reviews have five-star ratings, while only 2.5% have one-star ratings). We addressed this issue by treating the star rating as ordinal instead of metric and therefore estimating a proportional odds (PO) model (O’Connell 2006).

Third, as the online retailers represent different industries, our results may be affected by random industry effects. Therefore, to account for possible clustering of observations within industries, we used a cumulative link mixed model (CLMM; Christensen 2023), with a proportional odds logit link for a more robust assessment. The model included the respective categorization variable as a fixed effect and a random intercept for the industry.

## *Results*

*Dictionary-based analyses.* We began by comparing star ratings between written reviews that contain at least one of the delivery-related keywords (59.0%) to reviews that contain none. As is evident in Table WA4, at this highly aggregated level, the two groups show only minor differences in the distribution across the star rating categories. Neither a Mann-Whitney-U test ( $p = .790$ ) nor the proportional odds model (odds ratio [OR] = .992,  $p$

= .790) reaches statistical significance, even when we account for random industry effects in a mixed model (OR = 1.008,  $p = .796$ ). These results suggest a nuanced role of the moment of delivery. Therefore, we continued by examining the individual keywords used in the dictionary separately. This revealed a stark difference in rating scores between the 17,451 reviews that contain the word “liefer” and the 3,693 reviews that contain any other delivery-related keyword. Ratings for the former, as shown in Table WA4, are significantly higher than ratings for the 14,678 reviews without any keywords (Mann-Whitney-U:  $p < .001$ ; PO: OR = 1.264,  $p < .001$ ; CLMM: OR = 1.290,  $p < .001$ ). In contrast, the 3,693 ratings that contain any of the other keywords but not “liefer” exhibit significantly lower ratings than ratings without any keyword (Mann-Whitney-U:  $p < .001$ ; PO: OR = .610,  $p < .001$ ; CLMM: OR = .609,  $p < .001$ ). Additionally, the two keyword categories (i.e., with vs. without “liefer”) differ significantly from each other (Mann-Whitney-U:  $p < .001$ ; PO: OR = 1.865,  $p < .001$ ; CLMM: OR = 1.912,  $p < .001$ ).<sup>2</sup> Thus, in certain scenarios, the delivery experience prompts customers to evaluate an online retailer more positively, but in other instances, it is linked to very negative evaluations.

We further explored these findings by examining random review texts from the two focal categories, which revealed interesting insights. Review texts containing the keyword “liefer” refer to the delivery in a positive way, often highlighting the immediacy or short waiting time (e.g., “Orders are processed and delivered quickly.”; “very fast delivery”). In contrast, reviews containing any of the other keywords not only exhibit these positive evaluations but mention problems or unmet expectations (e.g., “delivery [keyword: *Zustellung*] sometimes takes too long”; “unfortunately, shipping [keyword: *Versand*] took a

---

<sup>2</sup> To account for the multiple group comparisons, we also compared the three categories in simultaneous estimations. A Kruskal-Wallis test reveals significant differences across groups ( $\chi^2 = 182.157$ ,  $df = 2$ ,  $p < .001$ ), with all pairwise contrasts also exhibiting  $p$ -values  $< .001$ ; similarly, PO and CLMM estimations show significant effects of the dummy-coded predictors for each of the keyword categories.

week”). These anecdotal accounts further support the notion that a more nuanced and context-sensitive approach such as an LLM-based analysis is warranted.

**Table WA4. Star Rating Distribution by Dictionary Keyword Categories**

| Star rating:                      | 1             | 2             | 3             | 4               | 5                 | Total            | Mean<br>(SD)     |
|-----------------------------------|---------------|---------------|---------------|-----------------|-------------------|------------------|------------------|
| No delivery-related keyword       | 3.9%<br>(569) | 1.0%<br>(138) | 1.6%<br>(238) | 7.3%<br>(1,071) | 86.3%<br>(12,662) | 100%<br>(14,678) | 4.711<br>(.871)  |
| Any delivery-related keyword      | 2.7%<br>(574) | 1.1%<br>(227) | 2.0%<br>(415) | 8.3%<br>(1,743) | 86.0%<br>(18,185) | 100%<br>(21,144) | 4.738<br>(.789)  |
| Keyword “liefer”                  | 2.2%<br>(376) | .9%<br>(153)  | 1.8%<br>(315) | 7.7%<br>(1,348) | 87.4%<br>(15,259) | 100%<br>(17,451) | 4.774<br>(.723)  |
| Any other keyword except “liefer” | 5.4%<br>(198) | 2.0%<br>(74)  | 2.7%<br>(100) | 10.7%<br>(395)  | 79.2%<br>(2,926)  | 100%<br>(3,693)  | 4.564<br>(1.031) |

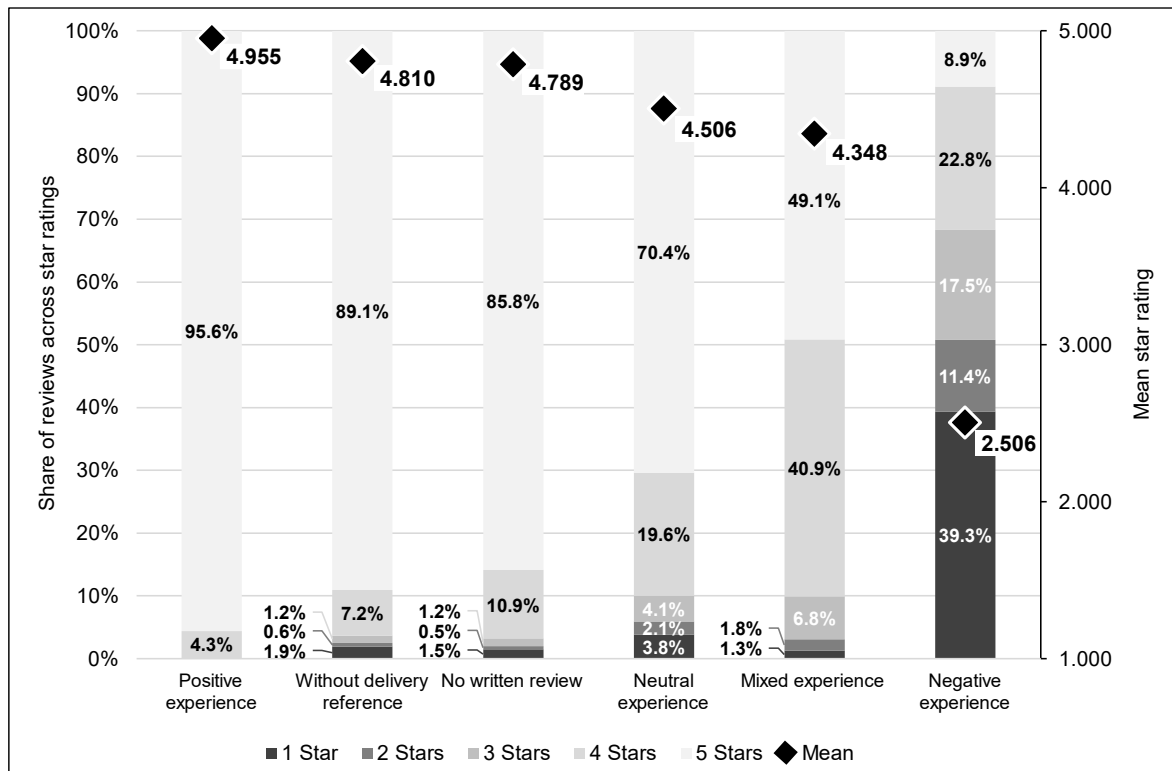
*LLM-based analyses.* Before analyzing the LLM-assisted categorizations, we compared the two approaches. This indicates a more detailed categorization based on the LLM compared to the dictionary approach, as expected. Of the 35,822 reviews with text, 2,747 (7.7%) of the reviews that do not contain any keyword were categorized by the LLM as referring to the delivery. In contrast, 289 (0.8%) of the reviews that do contain one of the keywords were categorized by the LLM as not referring to the delivery.

The comparisons in star ratings based on the categorizations provide support for our assumption that the moment of delivery is diagnostic of customers’ evaluations and WOM. First, as can be seen in Table WA5, ratings for the 23,602 reviews that refer to the delivery are significantly higher than ratings for the 12,220 reviews without any such reference (Mann-Whitney-U:  $p < .001$ ; PO: OR = .661,  $p < .001$ ; CLMM: OR = .674,  $p < .001$ ). Compared with our dictionary approach, the LLM-based categorization seemingly detects more delivery-related reviews, which are relevant for customer ratings, and thus offers a clearer distinction.

Second, we compared ratings across the different categorizations (see Table WA5 and Figure WA1); a Kruskal-Wallis test indicates significant differences ( $\chi^2 = 15,523.0$ ,  $df = 5$ ,  $p < .001$ ), and all pairwise contrasts exhibit  $p$ -values less than .001. Furthermore, both the PO ( $\chi^2 = 10,303.0$ ,  $df = 5$ ,  $p < .001$ ) and CLMM ( $\chi^2 = 9,907.99$ ,  $df = 5$ ,  $p < .001$ ) estimations indicate significant differences across categories. Reviews that praise the delivery, exhibit the highest star rating, followed by reviews without any text and reviews without any delivery reference; reviews that mention the delivery in a neutral way align with the next highest rating, followed by reviews with a mixed description of the delivery. Finally, reviews that refer to the delivery experience negatively are associated with the lowest ranking.

**Table WA5. Star Rating Distribution by LLM-Based Categories**

| Star rating:                  | 1              | 2              | 3              | 4                | 5                 | Total            | Mean<br>(SD)     |
|-------------------------------|----------------|----------------|----------------|------------------|-------------------|------------------|------------------|
| Without delivery<br>reference | 1.9%<br>(231)  | .6%<br>(75)    | 1.2%<br>(144)  | 7.2%<br>(885)    | 89.1%<br>(10,885) | 100%<br>(12,220) | 4.810<br>(.664)  |
| With delivery<br>reference    | 3.9%<br>(912)  | 1.2%<br>(290)  | 2.2%<br>(509)  | 8.2%<br>(1,929)  | 84.6%<br>(19,962) | 100%<br>(23,602) | 4.684<br>(.893)  |
| Neutral experience            | 3.8%<br>(16)   | 2.1%<br>(9)    | 4.1%<br>(17)   | 19.6%<br>(82)    | 70.4%<br>(295)    | 100%<br>(419)    | 4.506<br>(.959)  |
| Negative experience           | 39.3%<br>(878) | 11.4%<br>(255) | 17.5%<br>(391) | 22.8%<br>(510)   | 8.9%<br>(199)     | 100%<br>(2,233)  | 2.506<br>(1.424) |
| Positive experience           | .01%<br>(2)    | .02%<br>(4)    | .1%<br>(18)    | 4.3%<br>(839)    | 95.6%<br>(18,870) | 100%<br>(19,733) | 4.955<br>(.218)  |
| Mixed experience              | 1.3%<br>(16)   | 1.8%<br>(22)   | 6.8%<br>(83)   | 40.9%<br>(489)   | 49.1%<br>(598)    | 100%<br>(1,217)  | 4.348<br>(.794)  |
| No written review             | 1.5%<br>(360)  | .5%<br>(122)   | 1.2%<br>(289)  | 10.9%<br>(2,554) | 85.8%<br>(20,032) | 100%<br>(23,357) | 4.789<br>(.639)  |

**Figure WA1. Star Rating Distribution and Mean Rating by LLM-Based Categories**

### Discussion

Using field data involving more than 35,000 reviews of customers' experiences with online retailers, which they share with others, we test the base assumption that the moment of delivery represents a crucial determinant of customers' overall service evaluation and word-of-mouth activities. The analytical approach included two methods for identifying and categorizing customer accounts of the delivery experience—a more conventional dictionary-based method and an LLM-based categorization that exploits recent advances in AI. The latter provides more nuanced content analyses, but both approaches also provide consistent evidence of the importance of the delivery experience, across industries. Our analyses thus offer a clear indication that the delivery experience is diagnostic of how customers evaluate an online retailer.

## WEB APPENDIX B: STIMULUS MATERIAL (Studies 1a–1c)

### Introduction [Studies 1a–1c]

Please imagine the following situation:

Imagine that you have ordered a [book/backpack/pair of shoes] via the online retailer [\*name of the online retailer\*] You have already completed the ordering process. A few hours after placing your order, you receive the following shipping confirmation from [\*name of the online retailer\*]:

#### Shipping confirmation [Studies 1a–1c]

Hello,

We would like to inform you that we have now dispatched your order.

The shipment is carried out by our [delivery service provider \*name of the third-party provider\*/own logistics service provider \*name of the online retailer\* delivery]; it is no longer possible for you or our customer service to change the order. If you would like to return an item from your order or view or change other orders, you can easily do so via "My Orders" on our website.

### Study 1a: Moment of Delivery Experience (video<sup>a</sup>)

#### Negative



Third-Party (QuickShip)

[Watch video](#)



Proprietary (shoezone)

[Watch video](#)

#### Positive



Third-Party (QuickShip)

[Watch video](#)



Proprietary (shoezone)

[Watch video](#)

### Study 1b: Moment of Delivery Experience (dialogue)

#### Negative

One day after the announced delivery day, the [DHL/Hiking-World delivery] courier rings the doorbell. You open the door and an annoyed-looking courier dressed in a waistcoat stands in front of you. In the background you can see an old and dirty [DHL/Hiking-World] delivery van.

Imagine the following dialogue:

[DHL/Hiking-World] courier: "Hello, [DHL/Hiking-World], I have a parcel for you."

You: "Hello, great, I've been waiting for this parcel."

The annoyed courier scans the parcel with his handheld reader.

[DHL/Hiking-World] courier: "Now I need your signature."

You: "Gladly."

He hands you a sticky pen with which you sign directly on the device.

[DHL/Hiking-World] courier: "Your parcel, bye!"

You take the parcel.

You: "Thank you, bye."

The courier strolls demotivated back to his van. You close the door.

#### Positive

On the announced delivery day, the [DHL/Hiking-World delivery] courier rings the doorbell. You open the door and a friendly-looking courier in [DHL/Hiking-World] delivery clothing stands in front of you. In the background you can see a new and clean [DHL/Hiking-World] delivery van.

Imagine the following dialogue:

[DHL/Hiking-World] courier: "Good afternoon, [DHL/Hiking-World], I have a parcel for you."

You: "Hello, great, I've been waiting for this parcel."

The friendly courier scans the parcel with his handheld reader.

[DHL/Hiking-World] courier: "Now I need your signature, please."

You: "Gladly."

He hands you a clean pen with which you sign directly on the device.

[DHL/Hiking-World] courier: "Thank you, that's all. Here's your parcel. Have a nice day."

You take the parcel.

You: "Thank you, bye."

The courier walks motivated back to his van. You close the door.

### Study 1c: Moment of Delivery Experience (description)

#### Negative

One day after the announced delivery day, the [DHL/Amazon delivery] courier rings the doorbell. Without any greeting, the courier, dressed in a waistcoat, tries to scan the parcel with his handheld reader. After a few attempts, it worked. The courier hands you the parcel and leaves. You close the door and see that the parcel is damaged.

#### Positive

On the announced delivery day, the [DHL/Amazon delivery] courier rings the doorbell. You open the door and are greeted by a friendly courier dressed in a [DHL/Amazon delivery] uniform. The courier promptly scans the parcel with his handheld reader, hands it over to you, and wishes you a good day. You close the door and see that the parcel is undamaged.

<sup>a</sup> Links to the videos depicting the MoD experience are available by clicking on "Watch video".

## WEB APPENDIX C: STIMULUS MATERIAL (Study 1a, pretest)

### Mock-ups of both fictitious shoe retailers

| Focal retailer (shoezone) | Non-focal retailer (dreamshoes) |
|---------------------------|---------------------------------|
|                           |                                 |

## WEB APPENDIX D: ITEMS AND RELIABILITY MEASURES (Studies 1–2)

|                                                                                                 | Cronbach's $\alpha$ | Construct Reliability | Average Variance Extracted | Standardized Factor Loadings |
|-------------------------------------------------------------------------------------------------|---------------------|-----------------------|----------------------------|------------------------------|
| <b>Positive word of mouth<sup>a</sup></b><br>(Zeithaml et al. 1996)                             | .95/.93/.91         | .97/.96/.94           | .91/.88/.84                |                              |
| 1. I would say positive things about [*name of the online retailer*] to other people.           |                     |                       |                            | .95/.94/.91                  |
| 2. I would recommend [*name of the online retailer*] to someone who seeks my advice.            |                     |                       |                            | .97/.95/.92                  |
| 3. I would encourage friends and relatives to do business with [*name of the online retailer*]. |                     |                       |                            | .94/.93/.91                  |
| <b>Attitude toward (fictitious) retailer<sup>b</sup></b><br>(Spears and Singh 2004)             | .88/.93/.91/.92     | .93/.96/.93/.95       | .81/.88/.74/.87            |                              |
| 1. Unpleasant / pleasant                                                                        |                     |                       |                            | .90/.94/.92/.94              |
| 2. Unfavorable / favorable                                                                      |                     |                       |                            | .90/.92/.74/.92              |
| 3. Unlikable / likable                                                                          |                     |                       |                            | .91/.95/.84/.93              |
| 4. Unappealing / appealing                                                                      |                     |                       |                            | —/—/.89/—                    |
| 5. Bad / good                                                                                   |                     |                       |                            | —/—/.90/—                    |

<sup>a</sup> All three items are used in Studies 1b, 1c, and 2; Study 1a only includes the first item. Seven-point Likert-type scale anchored by strongly disagree (1) and strongly agree (7).

<sup>b</sup> All five items are used in Study 1c; Studies 1a, 1b, and 2 include only items one–three. Seven-point semantic differential scale.

## WEB APPENDIX E: MODERATED MEDIATION ANALYSIS (Study 1a)

### *Follow-up Analysis*

As is well-established in the service literature, the experienced quality of a service encounter strongly affects consumers' intentions and behaviors, including their PWOM (e.g., Gremler et al. 2001; Smith et al. 1999; Subramony et al. 2023). Such PWOM is also a key antecedent of downstream outcomes, such as sales performance and purchase behavior (Babić-Rosario et al. 2016; East et al. 2008). The results of Study 1a reveal a significant interaction between MoD experience and courier type on both PWOM measures, but no interaction effect on retailer choice. These findings suggest that the MoD experience  $\times$  courier type interaction may not directly affect outcome variables (e.g., retailer choice) but can shape intermediating variables such as PWOM. Based on the service literature and our results, we propose that PWOM mediates the effect of MoD experience on retailer choice and that the relationship between the MoD experience and PWOM is moderated by the courier type, motivating a follow-up analysis.

### *Method*

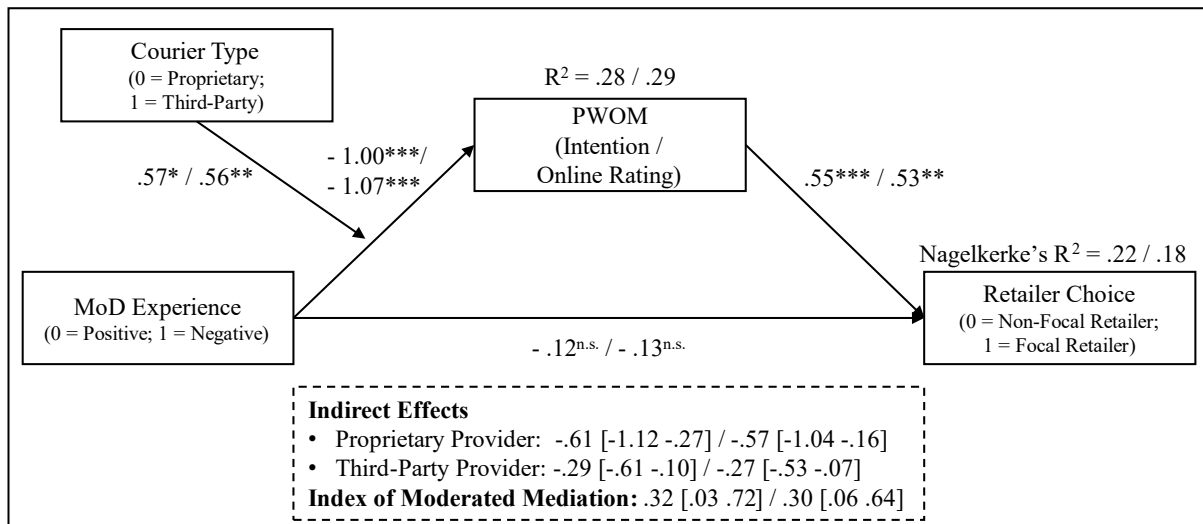
To examine potential moderated indirect effects, we perform a moderated mediation analysis in Hayes' (2022) PROCESS macro (Model 7; 5,000 bootstrap samples), using PWOM as mediator. To enhance the robustness of our results, we replicate the analysis with both PWOM measures (i.e., intention and online rating). Retailer choice is used as the binary dependent variable (0 = dreamshoes voucher / non-focal retailer; 1 = shoezone voucher / focal retailer), MoD experience (0 = positive; 1 = negative) was specified as the independent variable, and courier type (0 = proprietary; 1 = third-party) as the moderator. The same control variables used in prior analyses were included (age, gender, attitude toward the retailer, general online spending).

## Results

*PWOM intention.* According to the results in Figure WA2, the MoD experience exerts a significant, negative, indirect effect on retailer choice through PWOM intentions; the confidence interval (CI) around the estimate excludes 0. This indirect effect also is moderated by courier type, such that the negative indirect effect on retailer choice becomes mitigated if the delivery involves a third-party ( $b = -.29$ ,  $SE = .13$ , 95% CI:  $-.6120$ ,  $-.0951$ ) rather than proprietary ( $b = -.61$ ,  $SE = .22$ , 95% CI:  $-1.1193$ ,  $-.2662$ ; index of moderated mediation =  $.32$ , 95% CI:  $.0287$ ,  $.7203$ ) provider.

*PWOM behavior.* The same pattern emerges when the more behavioral PWOM measure, online rating, is considered as mediator (see Figure WA2). Again, the indirect effects are both negative, significant, and moderated by courier type (third party:  $b = -.27$ ,  $SE = .12$ , 95% CI:  $-.5319$ ,  $-.0695$ ; proprietary:  $b = -.57$ ,  $SE = .22$ , 95% CI:  $-1.0374$ ,  $-.1639$ ; index of moderated mediation =  $.30$ , 95% CI:  $.0572$ ,  $.6432$ ).

**Figure WA2. Direct and (Conditional) Indirect Effects of MoD Experience through PWOM on Retailer Choice by Courier Type**



Notes: Process Model 7. Values before the slash refer to the model with PWOM intention as mediator, while the values after the slash refer to the more behavior-related PWOM measure (online rating) as mediator. Non-significant covariates are gender, age, and general online spending (monthly); only initial attitude toward the fictitious online retailer indicates a significant effect on PWOM ( $b = .42 / .21$ ,  $SE = .06 / .04$ ,  $p < .001 / .001$ ) and retailer choice ( $b = .33 / .41$ ,  $SE = .13 / .12$ ,  $p < .05 / .001$ ).  $n.s.$   $p \geq .05$ ;  $*$   $p < .05$ ;  $**$   $p < .01$ ;  $***$   $p < .001$ .

*Discussion*

Our follow-up analysis reveals that the MoD experience indirectly affects retailer choice through PWOM. Moreover, this indirect effect is moderated by courier type, such that the negative indirect effect on retailer choice is attenuated when the delivery is handled by a third-party rather than a proprietary provider. These findings are robust and consistent across both PWOM measures, highlighting that courier type selection can be crucial, as it may directly influence intermediating variables (i.e., PWOM) and indirectly affect downstream outcomes such as retailer choice.

## WEB APPENDIX F: STIMULUS MATERIAL (Study 2)

### Introduction

Please imagine the following situation:

Imagine that you have ordered a new pair of shoes via the online retailer shoezone. You have already completed the ordering process. A few hours after placing your order, you receive the following shipping confirmation from shoezone:

### Shipping confirmation

Hello,

We would like to inform you that we have now dispatched your order.

The shipment is carried out by [delivery service provider DHL/our own logistics service provider shoezone delivery]; it is no longer possible for you or our customer service to change the order. If you would like to return an item from your order or view or change other orders, you can easily do so via "My Orders" on our website, shoezone.de.

### Inoculation<sup>a</sup>

Please note:

- Your satisfaction is our priority, but unfortunately, with millions of parcel deliveries every day, not all delivery errors can be avoided.
- We are in close contact with [the delivery service provider DHL/our delivery staff] to optimize processes and minimize potential errors.
- We also regularly review the dispatch processes to identify and rectify problems as quickly as possible.

We look forward to your next visit!

shoezone.de

### Moment of delivery (video <sup>b</sup>)

#### Negative



Third-Party  
(DHL)

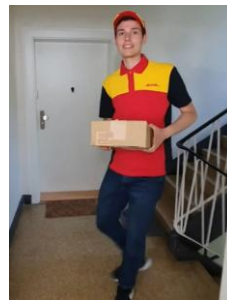
[Watch video](#)



Proprietary  
(shoezone)

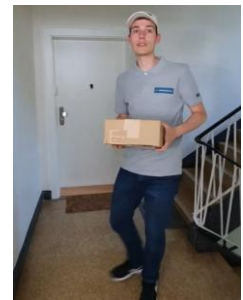
[Watch video](#)

#### Positive



Third-Party  
(DHL)

[Watch video](#)



Proprietary  
(shoezone)

[Watch video](#)

<sup>a</sup> Inoculation message only appeared in the inoculation-present conditions.

<sup>b</sup> Links to the videos depicting the MoD experience are available by clicking on "Watch video".

## REFERENCES

- Arlinghaus, Clarissa S., Charlotte Wulff and Günter W. Maier. 2024. "Inductive Coding with ChatGPT—An Evaluation of Different GPT Models Clustering Qualitative Data into Categories." *OSF Preprints*. [https://osf.io/preprints/osf/gpnyv\\_v1](https://osf.io/preprints/osf/gpnyv_v1)
- Babić-Rosario, Ana, Francesca Sotgiu, Kristine de Valck and Tammo H. Bijmolt. 2016. "The Effect of Electronic Word of Mouth on Sales: A Meta-Analytic Review of Platform, Product, and Metric Factors." *Journal of Marketing Research* 53, no. (3): 297-318.
- Christensen, Rune H. B. 2023. "Ordinal—Regression Models for Ordinal Data. R package version 2023.12-4.1." Retrieved May 18, 2026, from <https://CRAN.R-project.org/package=ordinal>
- East, Robert, Kathy Hammond and Wendy Lomax. 2008. "Measuring the Impact of Positive and Negative Word of Mouth on Brand Purchase Probability." *International Journal of Research in Marketing* 25, no. (3): 215-24.
- Gremler, Dwayne D., Kevin P. Gwinner and Stephen W. Brown. 2001. "Generating Positive Word-of-Mouth Communication Through Customer-employee Relationships." *International Journal of Service Industry Management* 12, no. (1): 44-59.
- Hayes, Andrew F. 2022. *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach*. 3rd ed. New York: Guilford Press.
- O'Connell, Ann. 2006. *Logistic Regression Models for Ordinal Response Variables*. Thousand Oaks, CA: SAGE Publications.
- Smith, Amy K., Ruth N. Bolton and Janet Wagner. 1999. "A Model of Customer Satisfaction with Service Encounters Involving Failure and Recovery." *Journal of Marketing Research* 36, no. (3): 356-72.
- Spears, Nancy and Surendra N. Singh. 2004. "Measuring Attitude Toward the Brand and Purchase Intentions." *Journal of Current Issues & Research in Advertising* 26, no. (2): 53-66.
- Subramony, Mahesh, Danielle van Jaarsveld, Helena Nguyen, Markus Groth, and David Solnet. 2023. "Frontline Service Employee Research: Integration of Systematic Literature Reviews and Recommendations for Future Scholarship." *Journal of Service Management* 34, no. (5): 829-42.
- Villarroel Ordenes, Francisco, Grant Packard, Jochen Hartmann, and Davide Proserpio. 2025. "Using Traditional Text Analysis and Large Language Models in Service Failure and Recovery." *Journal of Service Research* 28, no. (3): 382-88.
- Zeithaml, Valarie A., Leonard L. Berry, and A. Parasuraman. 1996. "The Behavioral Consequences of Service Quality." *Journal of Marketing* 60, no. (2): 31-46.