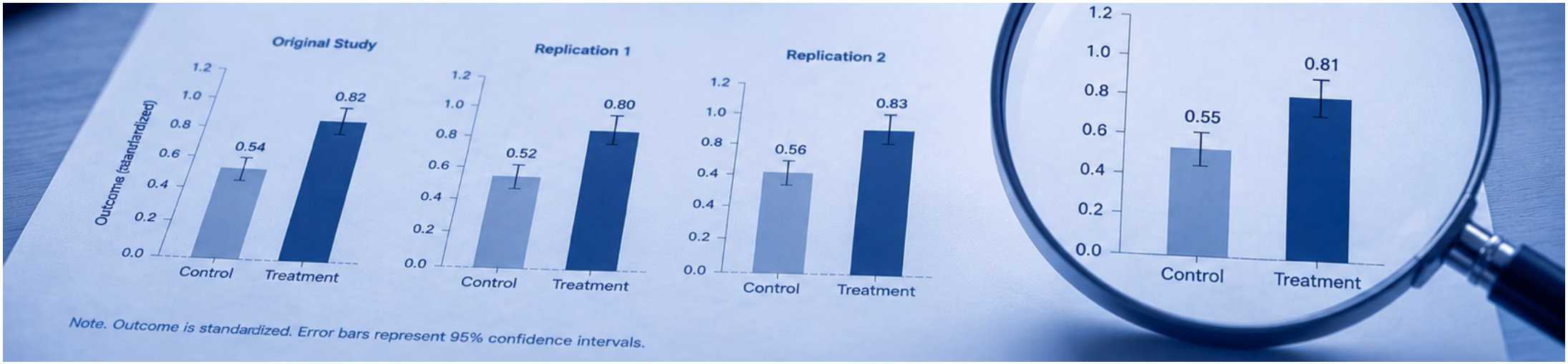
UNIVERSITÄT TRIER

Replikationen in der betriebswirtschaftlichen Forschung

Sören Köcher

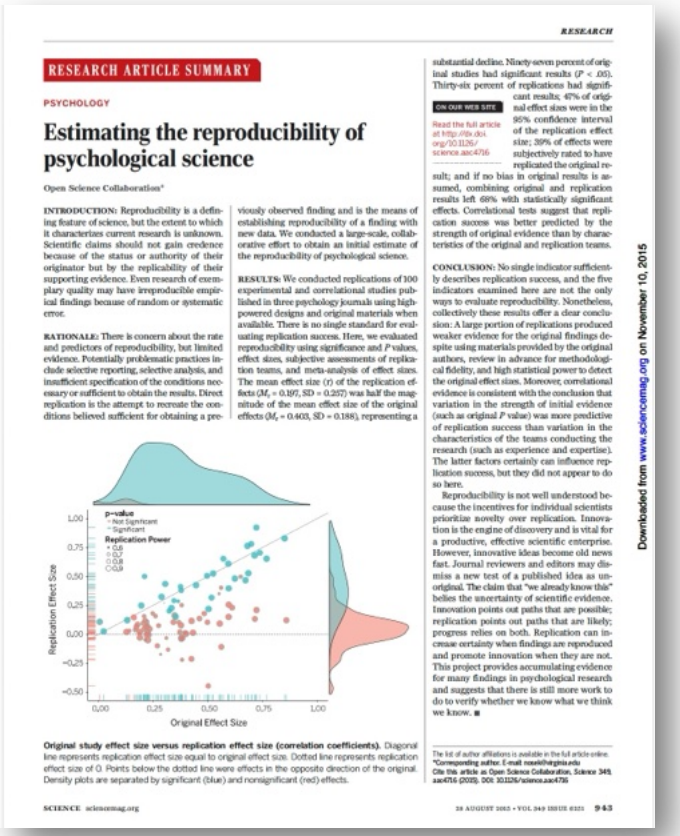


Tagung: Informationsversorgung für die Wirtschaftsforschung:
Zwischen Daten, Offenheit und KI
Kiel, 9. Juli 2026

MARKETING & INNOVATION

Die Replikationskrise

Was ist passiert?



RESEARCH

RESEARCH ARTICLE SUMMARY

PSYCHOLOGY

Estimating the reproducibility of psychological science

Open Science Collaboration*

INTRODUCTION: Reproducibility is a defining feature of science, but the extent to which it characterizes current research is unknown. Scientific claims should not gain credence because of the status or authority of their originator but by the replicability of their supporting evidence. Even research of exemplary quality may have irreproducible empirical findings because of random or systematic error.

RATIONALE: There is concern about the rate and predictors of reproducibility, but limited evidence. Potentially problematic practices include selective reporting, selective analysis, and insufficient specification of the conditions necessary or sufficient to obtain the results. Direct replication is the attempt to recreate the conditions believed sufficient for obtaining a previously observed finding and is the means of establishing reproducibility of a finding with new data. We conducted a large-scale, collaborative effort to obtain an initial estimate of the reproducibility of psychological science.

RESULTS: We conducted replications of 100 experimental and correlational studies published in three psychology journals using high-powered designs and original materials when available. There is no single standard for evaluating replication success. Here, we evaluated reproducibility using significance and *P* values, effect sizes, subjective assessments of replication teams, and meta-analysis of effect sizes. The mean effect size (*r*) of the replication effects (M_r = 0.197, SD = 0.257) was half the magnitude of the mean effect size of the original effects (M_r = 0.403, SD = 0.188), representing a substantial decline. Ninety-seven percent of original studies had significant results ($P < .05$). Thirty-six percent of replications had significant results; 47% of original effect sizes were in the 95% confidence interval of the replication effect size; 39% of effects were subjectively rated to have replicated the original result; and if no bias in original results is assumed, combining original and replication results left 68% with statistically significant effects. Correlational tests suggest that replication success was better predicted by the strength of original evidence than by characteristics of the original and replication teams.

CONCLUSION: No single indicator sufficiently describes replication success, and the five indicators examined here are not the only ways to evaluate reproducibility. Nonetheless, collectively these results offer a clear conclusion: A large portion of replications produced weaker evidence for the original findings despite using materials provided by the original authors, review in advance for methodological fidelity, and high statistical power to detect the original effect sizes. Moreover, correlational evidence is consistent with the conclusion that variation in the strength of initial evidence (such as original *P* value) was more predictive of replication success than variation in the characteristics of the teams conducting the research (such as experience and expertise). The latter factors certainly can influence replication success, but they did not appear to do so here.

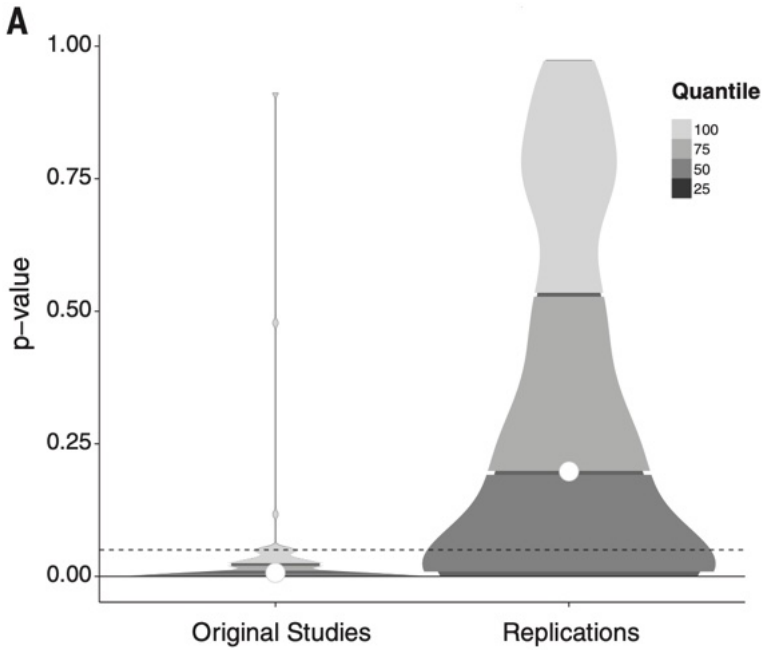
Reproducibility is not well understood because the incentives for individual scientists prioritize novelty over replication. Innovation is the engine of discovery and is vital for a productive, effective scientific enterprise. However, innovative ideas become old news fast. Journal reviewers and editors may dismiss a new test of a published idea as unoriginal. The claim that "we already know this" belies the uncertainty of scientific evidence. Innovation points out paths that are possible; replication points out paths that are likely; progress relies on both. Replication can increase certainty when findings are reproduced and promote innovation when they are not. This project provides accumulating evidence for many findings in psychological research and suggests that there is still more work to do to verify whether we know what we think we know. ■

ON OUR WEB SITE

Read the full article at http://dx.doi.org/10.1126/science.aac4716

The list of author affiliations is available in the full article online.
*Corresponding author. E-mail: nosek@virginia.edu
Cite this article as Open Science Collaboration, Science 349, aac4716 (2015). DOI: 10.1126/science.aac4716

Original study effect size versus replication effect size (correlation coefficients). Diagonal line represents replication effect size equal to original effect size. Dotted line represents replication effect size of 0. Points below the dotted line were effects in the opposite direction of the original. Density plots are separated by significant (blue) and nonsignificant (red) effects.



Ein Vergleich verschiedener Disziplinen

Article

Investigating the replicability of the social and behavioural sciences

https://doi.org/10.1038/s41586-025-10078-y
Received: 3 March 2025
Accepted: 19 December 2025
Published online: 1 April 2026
Check for updates

Pursuing replicability — independent evidence for previous claims — is important for creating generalizable knowledge[1,2]. Here we attempted replications of 274 claims of positive results from 164 quantitative papers published from 2009 to 2018 in 54 journals in the social and behavioural sciences. Replications were high powered on average to detect the original effect size (median of 99.6%), used original materials when relevant and available, and were peer reviewed in advance through a standardized internal protocol. Replications showed statistically significant results in the original pattern for 151 of 274 claims (55.1% (95% confidence interval (CI) 49.2–60.9%)) and for 80.8 of 164 papers (49.3% (95% CI 43.8–54.7%)), weighed for replicating multiple claims per paper. We observed modest variation in replication rates across disciplines (42.5–63.1%), although some estimates had high uncertainty. The median Pearson's *r* effect size was 0.25 (95% CI 0.21–0.27) for original studies and 0.10 (95% CI 0.09–0.13) for replication studies, an 82.4% (95% CI 67.8–88.2%) reduction in shared variance. Thirteen methods for evaluating replication success provided estimates ranging from 28.6% to 74.8% (median of 49.3%). Some decline in effect size and significance is expected based on power to detect original effects and regression to the mean because we replicated only positive results. We observe that challenges for replicability extend across social–behavioural sciences, illustrating the importance of identifying conditions that promote or inhibit replicability[3,4].

A central aim of science is to discover regularities in nature. If a claimed discovery is true, independent researchers should be able to conduct a similar investigation and reach similar conclusions. A replication attempt involves testing the same research question as a previous investigation with independent evidence, whether the evidence is a new data collection or existing secondary data that were not used in the previous investigation[1,2].

Across several previous replication studies in the social and behavioural sciences, totalling hundreds of replication attempts, approximately half of well-powered replication studies provided statistically significant evidence in the same direction as an original finding[3,5–12]. Moreover, observed effect sizes in replication studies were about half as large as effect sizes in original studies on average[3]. Similar results have been observed in other fields, such as preclinical cancer biology[13].

Popular explanations for the low observed replication success rates include underreporting of negative or inconclusive evidence for claims; high sampling error from small samples and measurement error due to unreliable measures, coupled with a statistical threshold ($P < 0.05$) that serves as a publication filter; low rigour and quality control in research design and measurement; and questionable research practices that inflate the likelihood of obtaining positive outcomes[3,14–20]. These factors are compounded by a research culture that rewards novel and 'interesting' findings and discourages error correction[21–23]. Replication failures are not necessarily due to the original findings having low credibility. Low replication rates can also be due to false negatives, poorly designed replications, selecting only positive results for replication, and differences between original and replication studies that are initially perceived as unimportant[2,24]. More broadly, there are conceptual challenges in deciding what is a replication of a previous finding and how to assess whether a replication attempt succeeded. Attempting replications provides a grounded context to wrestle with those challenges.

Identifying the potential causes of replication failures assumes that the existing evidence of replication failures is itself replicable. Most evidence comes from studies that examine replication outcomes within a single discipline and with relatively small samples. Here we report a systematic replication of quantitative published claims conducted as part of the Defense Advanced Research Projects Agency (DARPA) Systematizing Confidence in Open Research and Evidence (SCORE) program. Papers and claims were drawn from a sample of well-known journals selected to represent a diversity of subdisciplines across the social and behavioural sciences (Supplementary Table 1). The selected journals were aggregated for expository purposes into six disciplines: 11 journals in business (including organizational behaviour, management and marketing), 9 journals in economics (including finance), 7 journals in education, 7 journals in political science (including public administration), 13 journals in psychology (including health) and 7 journals in sociology (including criminology). See Supplementary Information for outcomes separately by subdiscipline, selection effects, effect size comparisons, outcomes across claims and subset analyses (Extended Data Figs. 1–6 and Supplementary Tables 8–19). For more details on the topics of the replication studies, see Extended Data Figs. 7 and 8 and Supplementary Table 20. Papers and claims eligible for inclusion had to be quantitative research of any kind and contain a statistical inference that identified a positive result using non-simulated human data including any level of human organization (for example, individuals, families, political entities, firms and economic units).

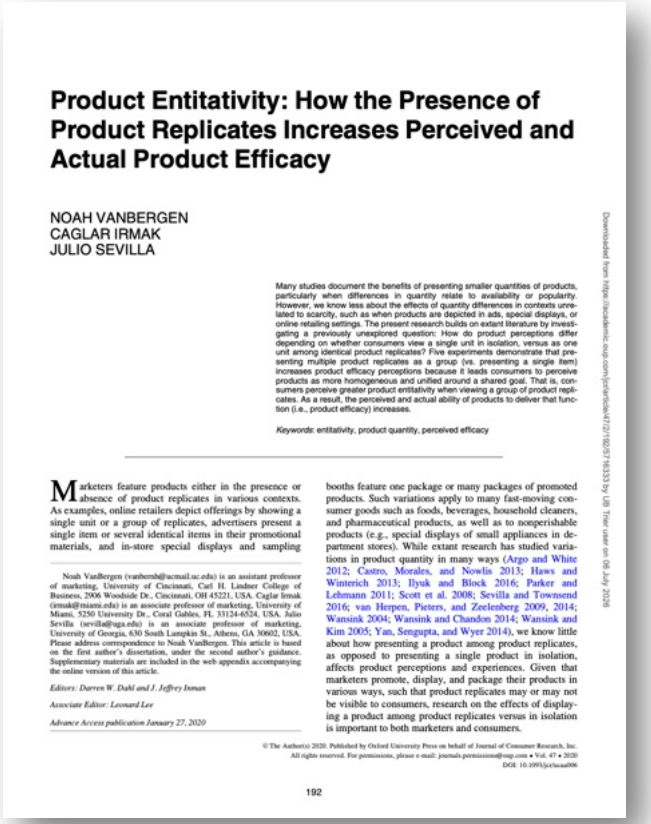
A list of authors and their affiliations appears at the end of the paper.

Table 3 | Original and replication outcomes by statistical significance and pattern and by Pearson's *r* effect size for six disciplines

Discipline	Replication attempt statistical significance and pattern		Median Pearson's *r* effect size (s.d.)	
	Counts	Percentage	Original estimate	Replication estimate
Business	17.0 of 36	47.2	0.24 (0.12)	0.10 (0.13)
Economics	10.2 of 24	42.5	0.28 (0.24)	0.13 (0.20)
Education	8.2 of 13	63.1	0.15 (0.28)	0.11 (0.10)
Political science	7.8 of 15	52.0	0.16 (0.22)	0.05 (0.16)
Psychology	28.4 of 58	49.0	0.29 (0.22)	0.11 (0.20)
Sociology	9.2 of 18	51.1	0.10 (0.16)	0.03 (0.17)

Papers are weighted combinations of claims accounting for multiple claims per paper replicated in some cases. The left column indicates the number of successful replications out of the total number of papers with replication attempts. Samples for the right columns are the papers for which a Pearson's *r* could be calculated for the original and replication outcomes.

Was ist eine gescheiterte Replikation?



Product Entitativity: How the Presence of Product Replicates Increases Perceived and Actual Product Efficacy

NOAH VANBERGEN
CAGLAR IRMAK
JULIO SEVILLA

Many studies document the benefits of presenting smaller quantities of products, particularly when differences in quantity relate to availability or popularity. However, we know less about the effects of quantity differences in contexts unrelated to scarcity, such as when products are depicted in ads, special displays, or online retailing settings. The present research builds on extant literature by investigating a previously unexplored question: How do product perceptions differ depending on whether consumers view a single unit in isolation, versus as one unit among identical product replicates? Five experiments demonstrate that presenting multiple product replicates as a group (vs. presenting a single item) increases product efficacy perceptions because it leads consumers to perceive products as more homogeneous and unified around a shared goal. That is, consumers perceive greater product entitativity when viewing a group of product replicates. As a result, the perceived and actual ability of products to deliver that function (i.e., product efficacy) increases.

Keywords: entitativity, product quantity, perceived efficacy

Marketers feature products either in the presence or absence of product replicates in various contexts. As examples, online retailers depict offerings by showing a single unit or a group of replicates, advertisers present a single item or several identical items in their promotional materials, and in-store special displays and sampling booths feature one package or many packages of promoted products. Such variations apply to many fast-moving consumer goods such as foods, beverages, household cleaners, and pharmaceutical products, as well as to nonperishable products (e.g., special displays of small appliances in department stores). While extant research has studied variations in product quantity in many ways (Argo and White 2012; Castro, Morales, and Nowlis 2013; Haws and Winterich 2013; Ilyuk and Block 2016; Parker and Lehmann 2011; Scott et al. 2008; Sevilla and Townsend 2016; van Herpen, Pieters, and Zeelenberg 2009, 2014; Wansink 2004; Wansink and Chandon 2014; Wansink and Kim 2005; Yan, Sengupta, and Wyer 2014), we know little about how presenting a product among product replicates, as opposed to presenting a single product in isolation, affects product perceptions and experiences. Given that marketers promote, display, and package their products in various ways, such that product replicates may or may not be visible to consumers, research on the effects of displaying a product among product replicates versus in isolation is important to both marketers and consumers.

Noah VanBergen (vanbernh@ucmail.uc.edu) is an assistant professor of marketing, University of Cincinnati, Carl H. Lindner College of Business, 2906 Woodside Dr., Cincinnati, OH 45221, USA. Caglar Irmak (irmak@miami.edu) is an associate professor of marketing, University of Miami, 5250 University Dr., Coral Gables, FL 33124-6524, USA. Julio Sevilla (sevilla@uga.edu) is an associate professor of marketing, University of Georgia, 630 South Lumpkin St., Athens, GA 30602, USA. Please address correspondence to Noah VanBergen. This article is based on the first author's dissertation, under the second author's guidance. Supplementary materials are included in the web appendix accompanying the online version of this article.

Editors: Darren W. Dahl and J. Jeffrey Inman

Associate Editor: Leonard Lee

Advance Access publication January 27, 2020

 • Vol. 47 • 2020
DOI: 10.1093/jcr/ucaa006

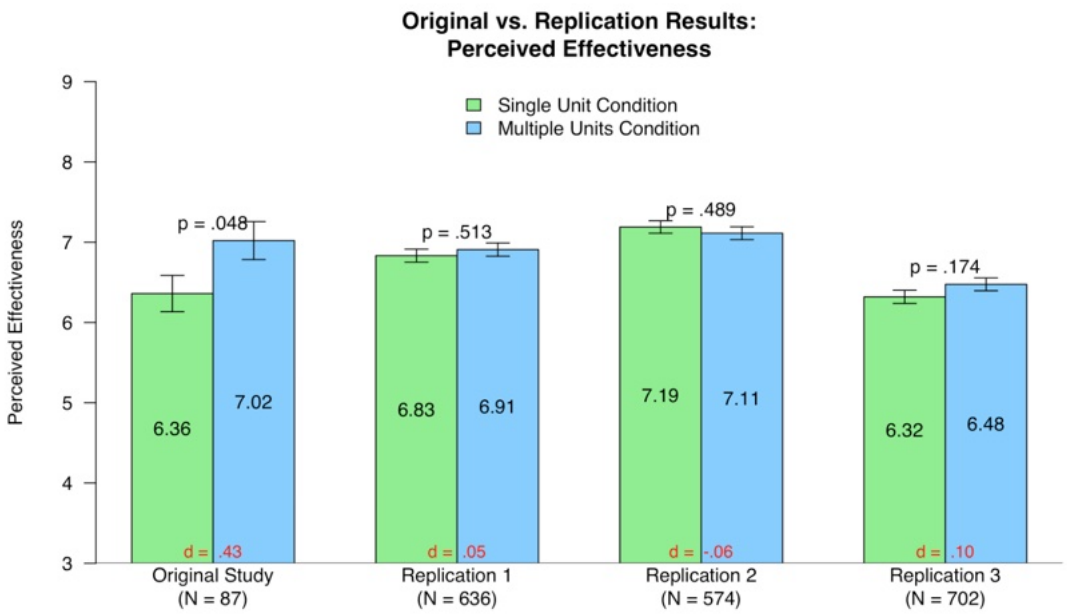
https://datacolada.org/90

Below is the mock-up advertisement we would like you to consider. Please take about 30 seconds to look over the ad. Try to think about the ad naturally, and on the next page we will ask you about what ideas came to mind based on the ad.

After viewing the ad, participants then proceeded to a separate page on which they answered four questions that measured how effective they believed Dettol to be. Those questions were:

1. In general, how effective do you think the Dettol you considered would be?
2. How much do you think Dettol would help you clean your home?
3. How effective do you think Dettol is in cleaning dirt and bacteria?
4. How potent do you think the active ingredients in Dettol are?

Was ist eine gescheiterte Replikation?



Mögliche Gründe

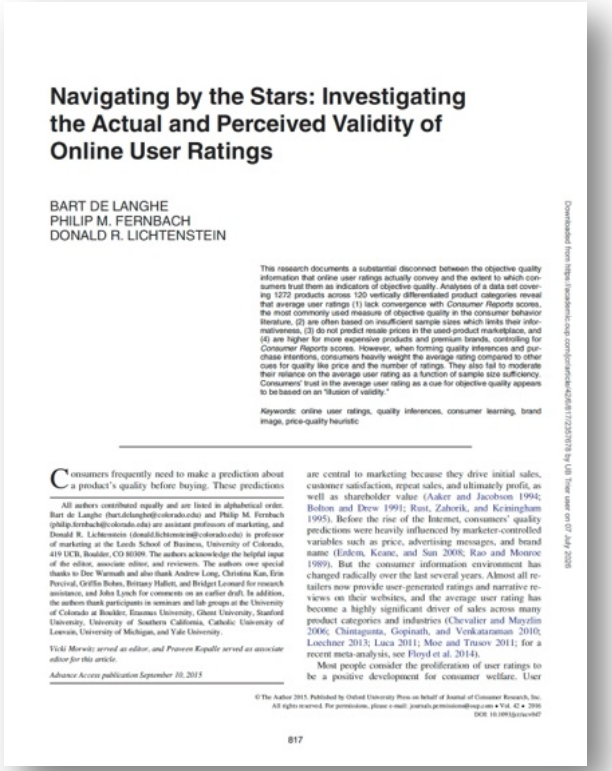
- Der Effekt existiert zwar, aber...
  - die Stichprobe unterscheidet sich,
  - der Effekt ist kleiner als ursprünglich angenommen.
- Der Effekt ist weniger robust als gedacht.
  - Er tritt nur unter bestimmten Bedingungen auf.
  - Es gibt bislang unbekannte Moderatoren.
- Der ursprüngliche Befund war überschätzt oder zufällig.
  - kleine Stichprobe,
  - geringe Power,
  - statistische Zufallsschwankung.
- Es gab Verzerrungen im Forschungsprozess.

https://datacolada.org/90

Arten von Replikationen

Replikationstyp	Ohne Erweiterung	Mit Erweiterung
Direct Replication Ist das Ergebnis reproduzierbar?	Exakte Wiederholung der Originalstudie mit identischem Design, identischen Materialien und derselben Stichprobe unter denselben Bedingungen (praktisch kaum realisierbar.)	
Close Replication Bleibt der Effekt unter ähnlichen Bedingungen bestehen?	Möglichst originalgetreue Wiederholung. Das Forschungsdesign, die Operationalisierung und die Analyse orientieren sich eng an der Originalstudie; notwendige Anpassungen (z. B. Sprache, Stichprobe oder Zeit) werden auf ein Minimum beschränkt.	Replikationen mit gezielter Erweiterung, z. B. zusätzliche Moderatoren, Mediatoren, weitere Stichproben oder Länder, zusätzliche abhängige Variablen
Conceptual Replication Ist die Theorie gültig?	Überprüfung derselben theoretischen Hypothese mit einer anderen Operationalisierung der unabhängigen und/oder abhängigen Variablen oder einem veränderten Studiendesign. Ziel ist die Prüfung der Generalisierbarkeit der Theorie, nicht des konkreten Experiments.	

Close Replication – Sollten wir nach den Sternen greifen?



Navigating by the Stars: Investigating the Actual and Perceived Validity of Online User Ratings

BART DE LANGHE
PHILIP M. FERNBACH
DONALD R. LICHTENSTEIN

This research documents a substantial disconnect between the objective quality information that online user ratings actually convey and the extent to which consumers trust them as indicators of objective quality. Analyses of a data set covering 1272 products across 120 vertically differentiated product categories reveal that average user ratings (1) lack convergence with *Consumer Reports* scores, the most commonly used measure of objective quality in the consumer behavior literature, (2) are often based on insufficient sample sizes which limits their informativeness, (3) do not predict resale prices in the used-product marketplace, and (4) are higher for more expensive products and premium brands, controlling for *Consumer Reports* scores. However, when forming quality inferences and purchase intentions, consumers heavily weight the average rating compared to other cues for quality like price and the number of ratings. They also fail to moderate their reliance on the average user rating as a function of sample size sufficiency. Consumers' trust in the average user rating as a cue for objective quality appears to be based on an "illusion of validity."

Keywords: online user ratings, quality inferences, consumer learning, brand image, price-quality heuristic

Consumers frequently need to make a prediction about a product's quality before buying. These predictions are central to marketing because they drive initial sales, customer satisfaction, repeat sales, and ultimately profit, as well as shareholder value (Aaker and Jacobson 1994; Bolton and Drew 1991; Rust, Zahorik, and Keiningham 1995). Before the rise of the Internet, consumers' quality predictions were heavily influenced by marketer-controlled variables such as price, advertising messages, and brand name (Erdem, Keane, and Sun 2008; Rao and Monroe 1989). But the consumer information environment has changed radically over the last several years. Almost all retailers now provide user-generated ratings and narrative reviews on their websites, and the average user rating has become a highly significant driver of sales across many product categories and industries (Chevalier and Mayzlin 2006; Chintagunta, Gopinath, and Venkataraman 2010; Loechner 2013; Luca 2011; Moe and Trusov 2011; for a recent meta-analysis, see Floyd et al. 2014).

Most people consider the proliferation of user ratings to be a positive development for consumer welfare. User

All authors contributed equally and are listed in alphabetical order. Bart de Langhe (bart.delanghe@colorado.edu) and Philip M. Fernbach (philip.fernbach@colorado.edu) are assistant professors of marketing, and Donald R. Lichtenstein (donald.lichtenstein@colorado.edu) is professor of marketing at the Leeds School of Business, University of Colorado, 419 UCB, Boulder, CO 80309. The authors acknowledge the helpful input of the editor, associate editor, and reviewers. The authors owe special thanks to Dee Warmath and also thank Andrew Long, Christina Kan, Erin Percival, Griffin Bolton, Brittany Hallett, and Bridget Leonard for research assistance, and John Lynch for comments on an earlier draft. In addition, the authors thank participants in seminars and lab groups at the University of Colorado at Boulder, Erasmus University, Ghent University, Stanford University, University of Southern California, Catholic University of Louvain, University of Michigan, and Yale University.

Vicki Morwitz served as editor, and Praveen Kopalle served as associate editor for this article.

Advance Access publication September 10, 2015

Vol. 42 • 2016
DOI: 10.1093/jcr/ucv047

- Online-Bewertungen korrelieren nur schwach mit den von *Consumer Reports* vergebenen Qualitätsbewertungen – die als Maß für die „objektive" Produktqualität angesehen werden können – korrelieren,
- Konsumentinnen und Konsumenten vertrauen bei der Beurteilung der Produktqualität stärker auf Online-Bewertungen, als es aufgrund deren tatsächlicher Aussagekraft gerechtfertigt wäre.

Amazonian Forests and Trees: Multiplicity and Objectivity in Studies of Online Consumer-Generated Ratings and Reviews, A Commentary on de Langhe, Fernbach, and Lichtenstein

ROBERT V. KOZINETS

Imperfect Progress: An Objective Quality Assessment of the Role of User Reviews in Consumer Decision Making, A Commentary on de Langhe, Fernbach, and Lichtenstein

ITAMAR SIMONSON

Star Wars: Response to Simonson, Winer/ Fader, and Kozinets

BART DE LANGHE
PHILIP M. FERNBACH
DONALD R. LICHTENSTEIN

Objective vs. Online Ratings: Are Low Correlations Unexpected and Does It Matter? A Commentary on de Langhe, Fernbach, and Lichtenstein

RUSSELL S. WINER
PETER S. FADER

Close Replication – Sollten wir nach den Sternen greifen?

Journal of Marketing Behavior, 2018, 3: 167–183

Replication Note

Should We Reach for the Stars? Examining the Convergence Between Online Product Ratings and Objective Product Quality and Their Impacts on Sales Performance

Sarah Köcher and Sören Köcher*
TU Dortmund University, Germany

ABSTRACT

By documenting that online ratings poorly correlate with quality scores provided by Consumer Reports — presumably a measure of "objective" product quality — de Langhe *et al.* (2016a) found that consumers rely more heavily on such ratings when making quality inferences than they should. Apart from replicating this finding, we examine the moderating effect of product age on the convergence between objective and rated quality and investigate which quality indicator is a better predictor of sales performance.

Keywords: Online product ratings; objective quality; sales performance; product age

Introduction

De Langhe *et al.* (2016a) recently discovered a substantial gap between the extent to which consumers trust in online product ratings when making

*Sarah Köcher, Doctoral Candidate in Marketing, Department of Marketing, TU Dortmund University; sarah.koecher@tu-dortmund.de. Sören Köcher, Assistant Professor of Marketing, Department of Marketing, TU Dortmund University, soeren.koecher@tu-dortmund.de. The two authors contributed equally to this paper. The authors thank the replication editor and the reviewers for their valuable feedback.

MS submitted on February 8, 2018; final version received May 15, 2018
Replication Editor: Eric Bradlow
ISSN 2326-568X; DOI 10.1561/107.00000050

Table 1: Parameter estimates (and confidence intervals) for the original and present study.

	de Langhe *et al.* (2016a)		Present study	
	Model A	Model B	Model A	Model B
Dependent variable	*Consumer Reports* quality scores		*Stiftung Warentest* quality scores	
Independent variables				
Average rating	0.16 (0.10 to 0.22)	0.09 (0.03 to 0.15)	0.13 (0.07 to 0.19)	0.08 (0.02 to 0.13)
Price		0.34 (0.28 to 0.39)		0.31 (0.26 to 0.36)
Standard error	−0.13 (−0.20 to −0.07)	−0.15 (−0.21 to −0.09)	−0.15 (−0.21 to −0.09)	−0.18 (−0.23 to −0.12)
Average rating × standard error	−0.06 (−0.12 to −0.01)	−0.07 (−0.12 to −0.02)	−0.01 (−0.06 to 0.04)	−0.01 (−0.05 to 0.04)
Average correlation between average rating and objective quality scores	0.18		0.18	
Percentage of negative correlations	34		36.3	
Data source of independent variables	Amazon.com		Amazon.de	
Number of observations	$N = 1{,}272$ products across 120 categories		$N = 1{,}322$ consumer electronic products across 224 categories	

Conceptual Replication – Differential Price Framing

Article

AMA AMERICAN MARKETING ASSOCIATION

When "More" Seems Like Less: Differential Price Framing Increases the Choice Share of Higher-Priced Options

Journal of Marketing Research
2019, Vol. 56(3) 816–841

DOI: 10.1177/0022243719851490
journals.sagepub.com/home/mrj
SAGE

Thomas Allard, David J. Hardisty, and Dale Griffin

Abstract

Four experiments supported by six supplemental studies show that premium but higher-priced products (e.g., direct flights, larger-capacity data storage devices) are more popular when the additional cost is made explicit using differential price framing (DPF; e.g., "for $20 more") rather than being left implicit, as in standard inclusive price framing (IPF; e.g., "for $60 total"). The DPF effect is driven by pricing focalism: relative to IPF, DPF creates a focus on the price difference, which, because it is smaller than the total price, leads to lower perceived expensiveness and thus greater choice share for the premium option. This price framing effect is robust to displaying the total cost of the purchase, bad deals, and easy-to-compute price differences, and it appears to be uniquely effective in pricing contexts. However, DPF effects are reduced among consumers who adopt a slow and effortful decision process. These findings have implications for research on price partitioning, the design of effective pricing strategy, the sources of expensiveness perceptions in the marketplace, and consumer welfare.

Keywords

attribute framing, expensiveness, heuristic processing, price partitioning, pricing strategy

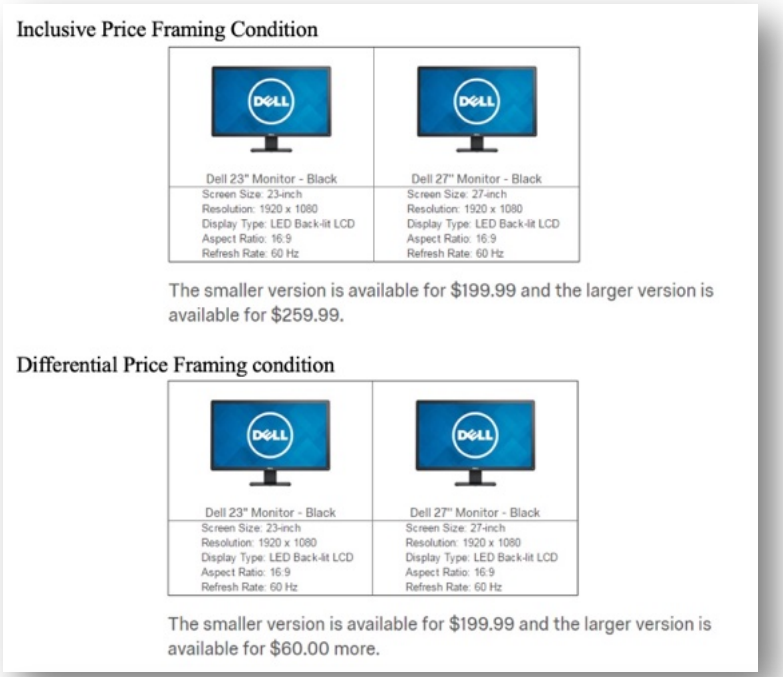
Online supplement: https://doi.org/10.1177/0022243719851490

Imagine shopping for a new computer monitor from a well-known brand. Deliberating between the standard 23-inch model and the more expensive 27-inch model, you notice that Retailer A quotes $199.99 for the 23-inch model and $259.99 for the 27-inch model. In contrast, retailer B quotes the prices as $199.99 for the 23-inch model and "$60.00 more" for the 27-inch model. How might these two pricing strategies influence your choices? Although simple arithmetic quickly reveals that both pricing formats lead to equivalent payments at the cash register, we predict that because of a representational bias caused by consumers focusing on the explicitly presented price format ("pricing focalism"), consumers will be more likely to purchase the 27-inch version at retailer B, where it is offered for "$60 more."

We test the notion that when selecting between a pair of products, with one being higher in quality and higher in price ("premium" option hereinafter), consumers are more likely to select the premium option when retailers quote the price *difference* from the standard option, rather than the *total* price of the premium option. Specifically, we define differential price framing (DPF), a price framing strategy that presents consumers with the upward price difference between vertically differentiated products (e.g., "$199.99" and "$60 more"). We contrast this approach with a more traditional pricing strategy—inclusive price framing (IPF)—that presents the total or aggregated price of each option (e.g., "$199.99" and "$259.99").

We show that DPF leads to stronger preferences for the premium product than IPF because focusing on the price difference, compared with focusing on the two total prices, leads consumers to perceive the premium option as less expensive. For instance, when selecting between a standard version of a product for X_1 and a premium quality version for X_2, framing the price of the premium option as a $\Delta X = (X_2 - X_1)$ increase will lead consumers to focus on the smaller ΔX amount (rather than X_2) as they make their subjective value assessment of the relative cost of the two products. Because $\Delta X << X_2$, and because most consumers use the information as presented to make expensiveness assessments (e.g., Raghubir and Srivastava 2002; Wertenbroch, Soman, and

Thomas Allard is Assistant Professor of Marketing, Nanyang Technological University, Singapore (email: tallard@ntu.edu.sg). David J. Hardisty is Assistant Professor of Marketing and Behavioural Science, University of British Columbia, Canada (email: david.hardisty@sauder.ubc.ca). Dale Griffin is Professor of Marketing and Behavioural Science, University of British Columbia, Canada (email: dale.griffin@sauder.ubc.ca).



Inclusive Price Framing Condition

Dell 23" Monitor - Black	Dell 27" Monitor - Black
Screen Size: 23-inch	Screen Size: 27-inch
Resolution: 1920 x 1080	Resolution: 1920 x 1080
Display Type: LED Back-lit LCD	Display Type: LED Back-lit LCD
Aspect Ratio: 16:9	Aspect Ratio: 16:9
Refresh Rate: 60 Hz	Refresh Rate: 60 Hz

The smaller version is available for $199.99 and the larger version is available for $259.99.

Differential Price Framing condition

Dell 23" Monitor - Black	Dell 27" Monitor - Black
Screen Size: 23-inch	Screen Size: 27-inch
Resolution: 1920 x 1080	Resolution: 1920 x 1080
Display Type: LED Back-lit LCD	Display Type: LED Back-lit LCD
Aspect Ratio: 16:9	Aspect Ratio: 16:9
Refresh Rate: 60 Hz	Refresh Rate: 60 Hz

The smaller version is available for $199.99 and the larger version is available for $60.00 more.

	IPF	DPF
% premium option	42%	58%

- In einer Reihe von zehn Labor- und Online-Studien zeigten Allard et al. (2019), dass Konsumentinnen und Konsumenten eine stärkere Präferenz für Premiumprodukte mit höherem Preis gegenüber Standardprodukten haben, wenn ihnen die zusätzlichen Kosten der teureren Option (z. B. „für 20 € mehr“) statt des Gesamtpreises („für insgesamt 60 €“) präsentiert werden
- Ob der Effekt sich auch in realem Einkaufsverhalten beobachten lässt ist jedoch noch unklar

Conceptual Replication – Differential Price Framing

Marketing Letters (2024) 35:159–170
https://doi.org/10.1007/s11002-023-09690-7

REPLICATION CORNER

A Conceptual replication of the differential price framing effect in the field

Sören Köcher[1] · Markus Husemann-Kopetzky[2] · Marie Schirmbeck[3] · Melina Hess[3] · Fabian Gmeindl[3] · Samuel Hess[3]

Accepted: 3 July 2023 / Published online: 12 July 2023

Abstract
Across a series of 10 laboratory and online studies, Allard, Hardisty, and Griffin (2019) demonstrated an increased preference for premium, higher-priced products over standard products when consumers were presented with the additional cost of the higher-priced option (i.e., differential price framing; e.g., "for $20 more") rather than with its total price (i.e., inclusive price framing; e.g., "for $60 total"); a phenomenon referred to as the differential price framing effect. In this paper, we conceptually replicate this effect in a field experiment that focuses on the application of a differential price framing strategy to a specific product format; namely, multipacks of identical products. Consistent with the differential price framing effect, the present study shows—based on 45,626 add-to-cart events and 30,426 completed product purchases on an online retailer's website—that the choice shares of higher-priced options increase when a differential price framing strategy is used. However, compared to non-consequential add-to-cart activities, this bias is considerably less pronounced in actual purchase patterns.

Keywords Differential Price Framing · Conceptual Replication · Partitioned Pricing · Consumer Choice · Field Experiment

✉ Sören Köcher
soeren.koecher@ovgu.de

[1] Otto-Von-Guericke-University Magdeburg, Faculty of Economics and Management, Chair of Marketing, Universitätsplatz 2, 39106 Magdeburg, Germany

[2] Freie Universität Berlin, School of Business and Economics, Marketing Department, Arnimallee 11, 14195 Berlin, Germany

[3] DRIP Agency, Kolpingstraße 32, 83278 Traunstein, Germany

Springer

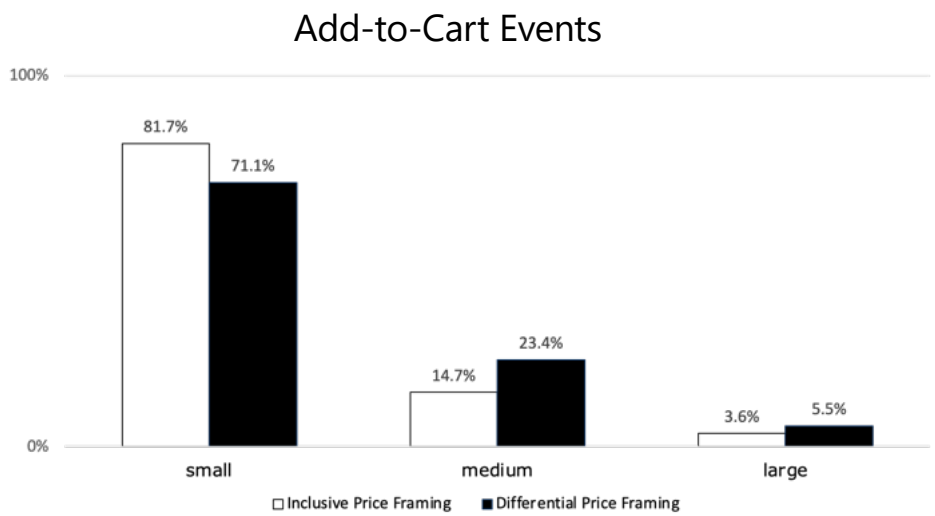
A) Inclusive price framing condition

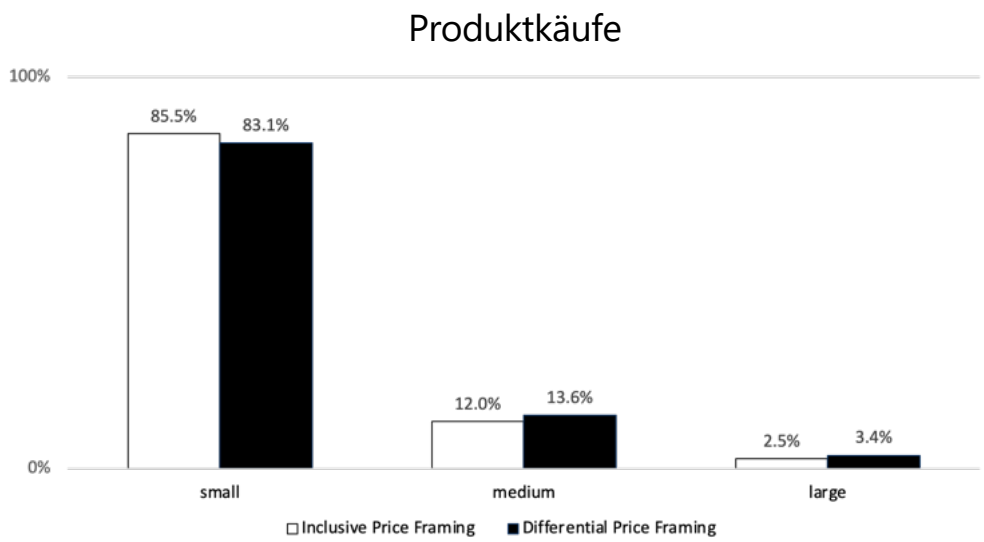
B) Differential price framing condition

Datengrundlage:

45.626 Add-to-Cart Events und 30.426 Produktkäufe
von 138.026 Websitebesuchern

Conceptual Replication – Differential Price Framing





- Kein Unterschied in den Conversion Rates, C_{IPF} = 11.8% vs. C_{DPF}= 11.7%
- Durchschnittlicher Bestellwert: M_{IPF} = 50.98 vs. M_{DPF} = 52.50 ($p < .001$)

Replikationen in der betriebswirtschaftlichen Forschung

Was Journals bevorzugen

- Neue Theorien
- Überraschende ("counterintuitive") Ergebnisse
- Hohe theoretische Originalität
- Signifikante Effekte

Neuheit ist ein zentrales Publikationskriterium

Replikationsstudien

- testen bestehende Erkenntnisse
- prüfen Robustheit und Generalisierbarkeit
- erweitern Theorien

häufig geringere Publikationschancen

Die Rolle von Open Science in der Replikationsforschung

Open-Science-Praktik	Originalstudien	Replikationsstudien
Open Materials	Materialien (z. B. Fragebögen, Stimuli, Instruktionen) offen bereitstellen, damit Studien möglichst originalgetreu repliziert werden können.	Verwendete Materialien und ggf. Anpassungen offen dokumentieren, damit weitere Replikationen möglich werden.
Open Data	Daten veröffentlichen, damit Ergebnisse überprüft, Effektgrößen nachvollzogen und Reanalysen durchgeführt werden können.	Daten veröffentlichen, damit Replikationsergebnisse überprüft, in Meta-Analysen integriert und weiterverwendet werden können.
Open Code	Analyseskripte bereitstellen, damit Ergebnisse reproduzierbar und nachvollziehbar sind.	Analyseskripte offenlegen, damit die Replikation vollständig reproduzierbar ist.
Präregistrierung	Hypothesen und Analysen vor Datenerhebung festlegen und transparent dokumentieren.	Replikationsdesign und Analyseplan vorab festlegen, um analytische Flexibilität zu reduzieren und die Glaubwürdigkeit zu erhöhen.
Open Access / Preprint	Studie frei zugänglich machen.	Replikationsergebnisse unabhängig vom Publikationserfolg zugänglich machen und so Publication Bias reduzieren.

Only when **certain events occur in accordance with rules or regularities**, as in the case of repeatable experiments, can our observations be tested – in principle – by anyone.

We do not take even our own observations quite seriously, or accept them as scientific observations, until we have repeated and tested them. Only by such repetitions can we convince ourselves that we are not dealing with a mere isolated 'coincidence,' but with events which, on account of their regularity and reproducibility, are in principle inter-subjectively testable.

(Popper 1959)

Herzlichen Dank für Ihre Aufmerksamkeit!

Prof. Dr. Sören Köcher
Professor of Marketing and Innovation

+49 651 201-3465
soeren.koecher@uni-trier.de
www.mui.uni-trier.de
linkedin.com/in/soeren-koecher-26777249