

Research Journal for Applied Management

Inhalt

de Alzaga Achter, Luisa; Khafif, Daniel-Ulf; Mühlbäck, Klaus:

KI-basierte Chatbots – eine kritische Analyse der Patientenerwartungen

Schuck, Katharina A.; Perret, Jens K.:

Optimizing Production of Fashion Goods as Means to a More Sustainable Garment Industry – An Assessment of the Literature

Gericke, Jens; Mehn, Audrey; Rommel, Kai:

Kundenpräferenzen für Bio-Lebensmittel in Deutschen Lebensmittelläden. Ergebnisse eines Discrete-Choice Experiments

Westermann, Arne; Böckenholt, Rebecca; Ghorbani, Mijka;

Messedat, Jons:

Interior Architectural Branding – The Implementation of the Brand Identity at the Point of Sale by Means of Interior Design

Kamran, Qeis; Becker, Marcus:

The Structure and Evolution of the Marketing Field: A Content Analysis of Five Decades of Research within the Academy of Marketing Science Journals

Becker, Marcus; Prokop Dayrell de Lima, Erika:

Natural Language Processing in IT Ticketing Systems. A conceptual framework for Question-and-Answering machines based on GPT-Algorithms

Research Journal for Applied Management

Jg. 4 (2023), Heft 1

ISM International School of Management
Prof. Dr. Ingo Böckenholt (Hg.)

Bibliografische Information der Deutschen Nationalbibliothek:
Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen
Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über
dnb.dnb.de abrufbar.



PubliQation – Wissenschaft veröffentlichen

Ein Imprint der Books on Demand GmbH, In de Tarpen 42, 22848 Norderstedt

© 2023 Ingo Böckenholt (Hg.)
ISM International School of Management gGmbH

Chefredaktion: Anette Herbst
Gestaltung: Anette Herbst, Jonas Weiler
Vorsitz Wissenschaftlicher Beirat: Prof. Dr. Kai Rommel

Herstellung und Verlag: BoD – Books on Demand GmbH
In de Tarpen 42, 22848 Norderstedt

ISBN Print	978-3-758-371-318
E-Book-ISBN	978-3-758-349-591

Inhalt

<i>Editorial</i>	III
de Alzaga Achter, Luisa; Khafif, Daniel-Ulf; Mühlbäck, Klaus	
<i>KI-basierte Chatbots – eine kritische Analyse der Patientenerwartungen</i>	1
Schuck, Katharina A.; Perret, Jens K.	
<i>Optimizing Production of Fashion Goods as Means to a More Sustainable Garment Industry – An Assessment of the Literature</i>	17
Gericke, Jens; Mehn, Audrey; Rommel, Kai	
<i>Kundenpräferenzen für Bio-Lebensmittel in deutschen Lebensmittelläden. Ergebnisse eines Discrete- Choice Experiments</i>	42
Westermann, Arne; Böckenholt, Rebecca; Ghorbani, Mijka; Messedat, Jons	
<i>Interior Architectural Branding – The Implementation of the Brand Identity at the Point of Sale by Means of Interior Design</i>	65
Kamran, Qeis; Becker, Marcus	
<i>The Structure and Evolution of the Marketing Field: A Content Analysis of Five Decades of Research within the Academy of Marketing Science Journals</i>	89
Becker, Marcus; Prokop Dayrell de Lima, Erika	
<i>Natural Language Processing in IT Ticketing Systems. A conceptual framework for Question-and- Answering machines based on GPT-Algorithms</i>	133

Editorial

The **Research Journal for Applied Management** (RJAM) offers current results from practice-oriented research on various management topics in issue 1/2024. The articles in this issue cover research topics in the areas of artificial intelligence, product optimization, consumer preferences, architectural branding and marketing research.

de Alzaga Achter, Khafif and Mühlbäck examine the expectations of patients in the German healthcare system towards AI-based chatbots and thus develop approaches to promote the use of such chatbots.

Schuck and Perret systematize existing production planning goals to show that production planning methods are extremely valuable tools for contributing to more sustainable clothing production.

Gericke, Mehn and Rommel examine the determinants of willingness to pay for organic products such as milk, meat, apples, potatoes and wine, which are permanently available in German markets for fast-moving consumer goods. The aim of the study is to determine the willingness to pay and to show the influence of selected socio-economic variables on the preference structure.

Westermann, Böckenholt, Ghorbani and Messedat develop an exploratory study design focused on an exemplary brand and examine how the brand personality is expressed in stores. The empirical study of the influence of store design on consumer perception of the brand shows that interior design elements are well suited to expressing the brand personality.

Kamran and Becker analyze five decades of research published in the “Journal of the Academy of Marketing Science” and “Academy of Marketing Science Review”. The authors use a Latent Dirichlet Allocation algorithm for unsupervised topic modeling. Using content analysis, they examine the relationships between marketing research streams that have developed over the four eras of the field's development.

Becker and Prokop present a study on the design of a Natural Language Processing (NLP) system within a Question-and-Answering (Q&A) environment for internal IT help desk ticketing operations. The proceedings will be used to develop a conversational agent for an IT consultancy company.

We would like to thank the reviewers of this issue and the Editorial Board for their assessment of the content of the submitted articles. We would also like to thank the team at the ISM Library and the staff of the Research Dean's Office for the successful implementation of the Research Journal for Applied Management. We hope all readers enjoy reading this issue. All issues of the Research Journal for Applied Management are available here: <https://ism.de/forschung/forschungsaktivitaeten> and <https://en.ism.de/university/research-activities>.

Editors: Prof. Dr. Ingo Böckenholt (President)

Editorial Management: Anette Herbst

Board of Editors: Dr. Markus B. Hofer (EbelHofer Strategy & Management Consultants GmbH), Prof. Dr. Ralf Kreutzer (HWR Berlin), Prof. Dr. Elmar Steurer (FH Neu-Ulm), Prof. Dr. Felix Streitferdt (TH Nürnberg), ISM: Prof. Dr. Manfred Bolin, Prof. Dr. Ernst Fahling, Prof. Dr. Silke Finken, Prof. Dr. Mike Hammes, Prof. Dr. Otto Jockel, Prof. Dr. Johannes Moskaliuk, Prof. Dr. Klaus Mühlbäck, Dr. Neumair, Prof. Dr. Eyden Samunderu, Prof. Dr. Dieter M. Schlesinger, Prof. Dr. Patrick Schmid, Prof. Dr. Michael Schmitt, Prof. Dr. Brigitte Spieß, Prof. Dr. Ernestine Schafmann.

Dortmund, December 2023

de Alzaga Achter, Luisa; Khafif, Daniel-Ulf; Mühlbäck, Klaus

KI-basierte Chatbots – eine kritische Analyse der Patientenerwartungen

Abstract

Ziel dieser Studie ist es, die Erwartungen von Patienten im deutschen Gesundheitswesen an KI-basierte Chatbots zu untersuchen und damit Ansatzpunkte zu entwickeln, die die Nutzung solcher Chatbots fördern können. Dazu werden KI-basierte Chatbots zunächst in die Rahmenbedingungen des deutschen Gesundheitsmarktes eingeordnet und ihre Bedeutung für die Zukunft der Patientenversorgung herausgearbeitet. Zur Beantwortung der Forschungsfrage wurde ein exploratives Forschungsdesign gewählt, bei dem acht Experten unterschiedlicher Alters- und Berufsgruppen mittels leitfadengestützter Interviews zu Erwartungen, Motivationen und Hemmnissen bei der Chatbotnutzung befragt wurden. Die Ergebnisse der Studie geben Hinweise zu Einsatzgebieten, Anforderungen sowie Grenzen und Problemen bei der Nutzung von Chatbots im Gesundheitswesen. Es zeigt sich, dass eine sich stetig verbessernde medizinische Kompetenz, d. h. zunehmende Intelligenz in der Verarbeitung auch komplexer Anamnesen, eine enge Interaktion mit dem behandelnden Arzt, hohe Anwendungsfreundlichkeit, verlässliche Funktionalität sowie Sicherheit der Datenspeicherung und -verarbeitung verbunden mit patientengerechter Sprache und Elementen menschlicher Empathie die Basis für zukünftiges Vertrauen von Patienten in KI-basierte Chatbots darstellen.

Keywords:

KI-basierte Chatbots, Gesundheitswesen, Patientenerwartungen

1 Einführung

Die digitale Transformation durchdringt unaufhaltsam alle Lebens- und Gesellschaftsbereiche und verändert dabei auch unser Gesundheitswesen grundlegend, indem sie neue Perspektiven für Diagnostik, Therapie und Prävention eröffnet. Dabei sind die Leistungserbringer im Gesundheitswesen mit dem demografischen Wandel konfrontiert, durch den die Anzahl an Patienten mit komplexen Krankheitsbildern steigt, während das Fachkräfteangebot schrumpft. Parallel dazu findet, auch angetrieben von Social Media- und E-Commerce-Trends, in der Bevölkerung ein Umdenken im Gesundheitsbewusstsein und Nachfrageverhalten statt. Die daraus resultierenden Erwartungen der Patienten erfordern von den Akteuren des Gesundheitswesens neue Ansätze, damit den Patienten eine angemessene und erwartungsgerechte Gesundheitsversorgung geboten werden kann. Zugleich steigt der Kosten- und Effizienzdruck auf die Leistungserbringer und Kostenträger. Traditionelle Branchengrenzen lösen sich auf und digitale Akteure wie E-Health-Startups drängen in die Gesundheitsbranche. (Laumer et al. 2019: 1).

In diesem Kontext gewinnt die Künstliche Intelligenz (KI) als Schlüsseltechnologie auch in der Medizin zunehmend an Bedeutung. Durch riesige Datenmengen und mithilfe von Algorithmen ist KI in der Lage, behandlungsrelevante Muster zu erkennen, Vorhersagen zu treffen sowie sich kontinuierlich anzupassen und zu lernen. Hierbei trifft KI eigenständig Entscheidungen und ist in der Lage, die menschliche Intelligenz zu ergänzen, zusätzliche Erkenntnisse zu liefern, Probleme zu lösen und die Produktivität zu verbessern (Zhang et al. 2020: 2; Nadarzynski et al. 2019: 2). Ein Ansatz, der zuletzt vermehrt Anwendung im Gesundheitswesen findet, ist der Einsatz von KI-basierten Chatbots.

Um die Akzeptanz von KI-basierten Chatbots in der Bevölkerung in Zukunft gewährleisten zu können, ist es für Anbieter und Entwickler von Chatbots von großer Bedeutung, die Erwartungen von Patienten an KI-basierte Chatbots zu kennen. Die noch geringe Nutzung von bereits bereitgestellten Chatbots durch Patienten impliziert, dass E-Health-Dienste bei Ärzten und Patienten bisher auf eher geringe Akzeptanz stoßen. Daher ist eine Erforschung der Motivation zur Nutzung von Gesundheits-Chatbots notwendig, um das Verhalten der Nutzer besser verstehen und Aspekte ableiten zu können, die bei der Entwicklung von Chatbots berücksichtigt werden sollten. Es liegen bislang nur wenige Studien vor, die die Akzeptanz von KI-basierten Chatbots bei den Nutzern untersucht haben (Nadarzynski et al. 2019: 1).

2 Theoretischer Rahmen & Stand der Forschung

Durch die zunehmende Digitalisierung des Gesundheitsmarkts hat sich das Rollenverständnis der Marktteilnehmer deutlich verändert. Auf der Seite der Leistungserbringer hat sich der Arzt vom einstigen ‚Halbgott in Weiß‘ zu einem Dienstleister entwickelt, der gemeinsam mit dem digital informierten und mündigen Patienten Entscheidungen trifft. Diese Entwicklung hat sich durch die Trans-

parenz von Gesundheitsinformationen und medizinischen Dienstleistungen, die die Selbstdiagnose und letztlich die Selbstoptimierung ermöglichen, vollzogen. Auf der Kostenträgerseite transformieren sich die Krankenkassen durch digitale Services vom ursprünglichen Prinzip als Kostenerstatter zum ganzheitlichen Gesundheitsanbieter. Innovationstreiber sind nun auch die Unternehmen der Tech-Start-up-Szene mit Gesundheits- und Medizin-Apps und nicht mehr nur die großen Unternehmen aus der Pharmaindustrie und Versicherungsbranche sowie die Politik (Seebach/Wasilewski 2021: 33). Aktuelle Errungenschaften der Digitalisierung und des Wandels im deutschen Gesundheitswesen sind zum Beispiel die elektronische Patientenakte, die Telemedizin, das elektronische Rezept oder KI-basierte Chatbots.

Chatbots, auch Conversational Agents (CA) genannt, sind Softwareprogramme, die über eine text- oder sprachbasierte Schnittstelle durch künstliche Kommunikation mit dem Nutzer interagieren.

Es gibt drei Formen, wie ein Chatbot implementiert werden kann: Chatbots können über Messaging-Apps, in eine Website integriert oder als Mobiltelefon-App eingesetzt werden. Konversationsagenten können dabei über eine Reihe von Methoden wie Text, Bild und Sprache mit dem Patienten kommunizieren, die menschliche Sprache interpretieren und mit synthetischen Stimmen antworten sowie vom Benutzer angeforderte Aufgaben erledigen (Laranjo et al. 2018: 1). Intelligente Gesprächsagenten haben das Potenzial, komplexere Aufgaben zu übernehmen, die mehr Interaktion, Schlussfolgerungen, Vorhersagen und Genauigkeit erfordern (Tudor Car et al. 2020: 2). Eine spezielle Form des Chatbots sind Embodied Conversational Agents (ECAs), Softwareeinheiten mit einer Verkörperung und Face-to-Face (F2F)-Interaktion (ter Stal et al. 2020: 1; Fitzpatrick et al. 2017: 2).

Chatsbots werden als nützlich diskutiert, um Barrieren zu überwinden und Patienten unabhängig von Ort und Zeit einen besseren Zugang zu Gesundheitsdiensten bei geringen Kosten zu ermöglichen. Während die ersten mobilen Health-Anwendungen (M-Health) nur einfache Funktionen boten ¹, können KI-basierte Chatbot-Systeme nicht nur als automatisierte Gesprächsagenten fungieren, sondern sind zudem in der Lage, die Gesundheit der Nutzer zu fördern, indem sie Diagnosevorschläge unterbreiten, als Informationsquelle dienen und so möglicherweise das Gesundheitsbewusstsein fördern (Laumer et al. 2019: 2). Chatbots sollen langfristig Anhaltspunkte bereitstellen können, um einen für alle Patienten gleichermaßen zugänglichen Ansatz für personalisierte medizinische Behandlungen und ein allgemein zugängliches evidenzbasiertes Wissen zu generieren (Liebrich 2017: 1).

In der Literatur besteht Einigkeit darüber, dass Patienten nicht passive Empfänger von Pflege sein sollen, sondern ‚aktiv‘ und ‚engagiert‘ an ihrer Pflege mitwirken sollen, um einen positiven Zugang zum Umgang mit Krankheiten und deren Behandlung zu finden (Pravettoni/Triberti 2020: 3). KI-basierte Chatbots ermöglichen es Patienten, spezifische Gesundheitsfragen zu klären, persönliche gesundheitsbezogene Erinnerungen einzurichten, mit klinischen Teams zu kommunizieren, Termine zu buchen,

¹ beispielsweise Apps zum Kalorienzählen

Gesundheitsdaten abzurufen und zu analysieren oder Diagnosemuster unter Berücksichtigung von Verhaltensindikatoren wie körperlicher Aktivität, Schlaf oder Ernährung zu übersetzen (Nadarzynski et al. 2019: 2).

Neben der Anwendung in der Therapie chronischer Krankheiten werden Chatbots gegenwärtig in der Gesundheitserziehung, Diagnostik und zur Begleitung und Unterstützung der psychischen Gesundheit eingesetzt. Chatbots können beispielsweise Antworten auf gesundheitsbezogene Anfragen von Patienten geben und gleichzeitig nach bestimmten Mustern von Symptomen nach Krankheiten suchen. Im Idealfall sind Chatbots in der Lage, Krankheiten im Frühstadium erkennen (Abashev et al. 2017: 1).

Durch den Einsatz von Chatbots im Gesundheitswesen ergeben sich für Patienten grundsätzlich Vorteile für ihre körperliche, psychologische und verhaltensbezogene Gesundheit (Bendig et al. 2019: 1; Palanica et al. 2019: 1). Zum Beispiel zeigen die Ergebnisse einer Einjahresstudie von Wefight Inc., bei der Brustkrebs-Patientinnen von einer Chatbot-Software begleitet und unterstützt wurden, dass die Patientinnen ihre Therapietreue und die Einhaltung der Medikamentengabe verbesserten (Chaix et al. 2019: 5f.).

Die Mehrheit der Patienten in Deutschland steht den digitalen Möglichkeiten im Gesundheitssystem offen gegenüber. In einer Online-Befragung von McKinsey im August 2020 gab jeder Dritte an, schon einmal einen Arztbesuch online vereinbart zu haben. Zwei von drei Deutschen begrüßen die Einführung der elektronischen Patientenakte und des elektronischen Rezepts – darunter mehr als 60 Prozent der Menschen in der Altersgruppe 65 Jahre und älter (Richter/Silberzahn 2020: 3).

Eine Studie aus den Niederlanden im Jahr 2010 mit ca. 900 Patienten zeigte eine hohe Akzeptanz von Chatbots, die über gesundheitsbezogene Themen aufklärten. Die Befragten betonten insbesondere die Anonymität der Kommunikation sowie die Qualität und Geschwindigkeit des Informationserhalts (Crutzen et al. 2011: 1). Ergebnisse einer Studie der Universität Osnabrück aus dem Jahr 2019 zeigen die Anwendbarkeit und das Potenzial von Chatbot-Systemen im Bereich der Verbesserung des Gesundheitsbewusstseins, insbesondere als neuartige Möglichkeit zur Entwicklung therapeutischer und präventiver Interventionen (Meier et al. 2019: 2). Eine schwedische Pilotstudie aus dem Jahr 2017, die einen vollautomatischen Chatbot zur Förderung des psychischen Wohlbefindens testete, hat die Wirksamkeit des Chatbots auf der Grundlage von kognitiver Verhaltenstherapie und Interaktionen der positiven Psychologie in der nicht-klinischen Bevölkerung nachgewiesen. Es zeigte sich ein signifikanter positiver Einfluss auf das Wohlbefinden und den empfundenen Stress, wobei einige Teilnehmer von einer besonderen 'digitalen Beziehung' zu dem Chatbot berichteten (Ly et al. 2017: 39; Nadarzynski et al. 2019: 2).

Obwohl die Technologie ständig weiterentwickelt wird, verfügen Chatbots derzeit nicht über vollständige Sprachfähigkeiten auf menschlichem Niveau, was zu Missverständnissen und Unzufriedenheit der Benutzer führen kann (Zhang et al. 2020: 2). Bei der Entwicklung von Algorithmen wird es für das maschinelle Lernen immer schwieriger, die Verarbeitung der Daten sowie die Gründe für die Antworten der Chatbots zu verfolgen. Dies ist als sog. Blackbox-Effekt bekannt (Tudor Car et al. 2020: 2).

Einer der Hauptkritikpunkte an CAs ist, dass sie kaum in der Lage sind, Empathie zu empfinden. Empathie bei einem Chatbot bedeutet, dass dieser die emotionalen Zustände der Benutzer erkennt und Antworten entsprechend dieser Emotionen findet. Ein Mangel an Empathie kann das Engagement, mit Chatbots zu kommunizieren, beeinträchtigen (Espinoza et al. 2020: 4). Chatbots, die eine Fähigkeit besitzen, Empathie zu erzeugen und auf Emotionen zu reagieren, führen zu einer höheren Compliance mit therapeutischen Behandlungen (Holmes et al. 2019).

Eine Hürde stellt für Health-Tech-Unternehmen die Befürchtung der Patienten vor Datenmissbrauch dar (Seebach/Wasilewski 2021: 350). Ein weiteres Hemmnis der Nutzung ist die geringe Qualität der derzeit verwendeten E-Health-Angebote. Zu den Problemen dieser Angebote gehören die geringe Leistung und die schlechte Nutzbarkeit, die es schwierig machen, die spezifischen Bedürfnisse der verschiedenen Nutzer zu erfüllen (Dubey/Rana 2010: 4723). Ein anderer Aspekt ist, dass zwar bereits ein großer Teil der medizinischen Einrichtungen in Deutschland eine hochdigitale Datenübertragung nutzt, es allerdings auch medizinische Einrichtungen gibt, die nach wie vor auf papierbasierte Prozesse angewiesen sind. Dies hat zur Konsequenz, dass eine Vernetzung und ein Austausch von Informationen zwischen den Leistungserbringern nicht immer stattfinden kann und daher keine reibungslose Datenübertragung gewährleistet ist (Liebrich 2017: 10).

Zusammenfassend korrespondieren die Ergebnisse der betrachteten Studien mit der allgemeinen Erkenntnis, dass E-Health-Dienste bei Ärzten und Patienten grundsätzlich auf Offenheit stoßen, während die Akzeptanz für Chatbots im Gesundheitswesen trotz des oben erwähnten breiten Spektrums an Anwendungsmöglichkeiten und eines zunehmenden Einsatzes von Chatbots in anderen Branchen noch relativ gering ist (Laranjo et al. 2018: 1255). Auch bestehen bei Ärzten Vorbehalte gegenüber KI-basierten Chatbots. Sie sind oft der Auffassung, dass eine Diagnose erst aufgrund eines Gesamteindrucks des Patienten und einer eingehenden Untersuchung möglich und eine persönliche Zuwendung des Arztes beim Arztbesuch für viele Patienten wichtig sei. Es ist auch zu bedenken, dass Ärzte durch den Einsatz von Chatbots Patienten verlieren könnten (Seebach/Wasilewski 2021: 97). Umgekehrt sind die gesetzlichen und privaten Krankenversicherungen als Kostenträger am Einsatz von Chatbots interessiert, um die Zahl der Arztbesuche und damit die Kosten zu senken. Auch der Staat hat Interesse an der Nutzung von Chatbots, da aufgrund des Ärztemangels in der Zukunft die Versorgung der Patienten gefährdet sein kann.

3 Ziele & methodisches Vorgehen

Derzeit gibt es lediglich einige wenige Studien zur Akzeptanz von Gesundheitschatbots und den Motiven der Nutzung durch die Patienten. Ein patientenzentrierter Ansatz erfordert die Untersuchung der Erwartungen von Patienten an KI-basierte Chatbots, aber auch potentieller Hemmnisse der Nutzung. Das Ziel dieser Studie ist es, Erwartungen von Patienten an KI-basierte Chatbots im deutschen Gesundheitswesen zu untersuchen und hieraus Ansatzpunkte zu generieren, die bei Entwicklung und Planung eines Chatbots berücksichtigt werden sollten. Zur Beantwortung dieser Leitfragen werden vier Teilaspekte untersucht:

- Wie groß ist die Hemmschwelle von Patienten bei der Nutzung KI-basierter Chatbots?
- Welche Motivationen haben Patienten, einen KI-basierten Chatbot zu nutzen, unter welchen Voraussetzungen tun sie dies?
- Welche Erwartungen haben Patienten an einen KI-basierten Chatbot?
- Welche Faktoren halten Patienten von der Chatbotnutzung ab?

Da die Auseinandersetzung mit Erwartungen von Patienten an KI-basierte Chatbots im deutschen Gesundheitswesen in der wissenschaftlichen Literatur bisher kaum Erkenntnisse geliefert hat, wurde für diese Studie ein exploratives Forschungsdesign auf Basis von insgesamt acht Experteninterviews gewählt.

Die Experteninterviews wurden einmalig im Sinne einer Querschnittsstudie durchgeführt. Grundlage für die Befragung war ein semistrukturierter Interviewleitfaden, der als Navigationshilfe für die Interviewer diente. Der Interviewleitfaden enthält nach einer kurzen Einführung in die Thematik sieben offene Kernfragen, die jeweils in mehrere Teilaspekte untergliedert sind, wodurch die definierten Forschungsziele dieser Studie vollständig abgedeckt sind, sowie einige Fragen zur Soziodemographie.

Um die unterschiedlichen Perspektiven verschiedener Anbieter und Nutzer von KI-basierten Chatbots zu untersuchen, wurden Experten aus verschiedenen Bereichen befragt: Entwickler von solchen Chatbots, Vertreter des Gesundheitswesens sowie Patienten, also Nutzer dieser Chatbots. Bei der Auswahl der Patienten wurde besonderer Wert darauf gelegt, Personen deutlich unterschiedlicher Soziodemographie zu wählen, um so redundanzfreie und breit gefächerte Sichtweisen zu generieren. Tabelle 1 listet die befragten Experten auf:

Tabelle 1: Übersicht über die befragten Experten

Codierung	Alter	Geschlecht	Beruf	Expertenkategorie
B1	78 J.	W	Rentnerin	Patient
B2	22 J.	M	Medizinstudent	Patient / Gesundheitswesen
B3	57 J.	M	Unternehmer	Patient
B4	23 J.	W	Studentin (nicht Medizin)	Patient
B5	26. J	W	E-Health-Produktmanagerin	Gesundheitswesen
B6	30 J.	M	E-Health-Softwareentwickler	Entwickler
B7	47 J.	M	Ärztlicher Direktor	Gesundheitswesen
B8	35 J.	M	CEO eines E-Health-Start-Ups	Gesundheitswesen / Entwickler

Quelle: eigene Darstellung

Die Gesprächsdauer mit den ausgewählten Experten betrug im Durchschnitt ca. 45 Minuten, wobei die Gespräche pandemiebedingt mittels des Online-Videokonferenzsystems Zoom durchgeführt und elektronisch dokumentiert wurden. Zur Auswertung der erhobenen Daten wurden die Gespräche mit Hilfe der Software Amberscript transkribiert, mit Absatz und Zeilennummerierungen versehen.

Die anschließende qualitative Inhaltsanalyse wurde nach Mayring (2015: 17) durchgeführt, wobei die Antworten als codierte Textpassagen kategorisiert wurden. Die systematische Auswertung der Antwortkategorien erfolgte mit Hilfe der Software MAXQDA. Generell wurde bei der Durchführung der Studie Wert darauf gelegt, den Anforderungen nach Validität, Objektivität und Reliabilität zu genügen.

4 Analyse, Ergebnisse und Interpretationen

Die Auswertung der Ergebnisse erfolgt kategorienbasiert, indem die Ergebnisse den vier Kernfragen zugeordnet werden.

4.1 Ergebnisse zur Hemmschwelle der Patienten

Mit einer Hemmschwelle wird in der Psychologie die Hinderung an einer Handlung beschrieben, wenn diese gegen die eigenen Gewohnheiten, Verhaltensweisen oder ähnliches stößt (Dudenredaktion (Bibliographisches Institut) 2015: 824). Aufgrund der breiten Alters- und Berufsverteilung der Experten ergaben sich breitgefächerte Einschätzungen zu den Hemmschwellen von Patienten. Es hat sich gezeigt, dass bei älteren Patienten eher eine hohe Hemmschwelle angenommen wird, wohingegen die Hemmschwelle bei jüngeren Patienten sowie bei Personen, die beruflich mit dem Gesundheitswesen oder dem IT-Bereich zu tun haben, als eher niedrig eingeschätzt wird. Mithin gehen die Experten davon aus, dass die Höhe der Hemmschwelle, KI-basierte Chatbots im Gesundheitswesen zu nutzen, stark

vom individuellen Patiententyp abhängt. Es wurde die Vermutung geäußert, dass eine hohe Hemmschwelle wesentlich damit zu begründen ist, dass die Chatbotnutzung für ältere Patienten die letzte Alternative zum persönlichen Arztbesuch sei. Für jüngere und im IT-Bereich tätige Personen hingegen gehöre es zum alltäglichen Verhalten zu chatten.

4.2 Ergebnisse zur Motivation zur Chatbotnutzung

Zunächst wurde durch die Experten als zentrale Motivation, KI-basierte Chatbots zu nutzen, der Aspekt Zeitersparnis genannt. Vor allem Menschen, die stark im beruflichen bzw. persönlichen Umfeld eingebunden sind, schätzen die Möglichkeiten eines Chatbots. Lange Wartezeiten in Arztpraxen bzw. häufige Telefonate für Termingesuche könnten dadurch vermieden werden. Die Experten führten aus, dass es als professionell empfunden werden könnte, wenn vor einem etwaigen Arztbesuch die Kommunikation mit einem Chatbot stehen würde. Dies entspräche einem Win-Win-Effekt im Sinne der Leistungsorientierung des Gesundheitssystems, da der vorbereitende Austausch durch den Chatbot-Einsatz sowohl dem Gesundheitssystem als auch dem Patienten eine Zeitersparnis brächte.

In diesem Zusammenhang wurde auch auf den Faktor Erreichbarkeit hingewiesen. Ein Chatbot ist im Unterschied zu den begrenzten Öffnungszeiten einer Arztpraxis rund um die Uhr erreichbar sowie zeitgleich beliebig oft reproduzierbar. Ein Anliegen, das der Patient am Abend dem Chatbot kommuniziert hat, kann am nächsten Morgen ggf. sofort dem Arzt mitgeteilt werden, so dass das Anliegen in der Sprechstunde nicht erneut geschildert werden muss.

Die Anonymität bei der Chatbotnutzung stellt nach Einschätzung der Experten eine weitere Motivation für den Fall dar, dass der Patient sich nicht wohl fühlt, dem Arzt gewisse Gesundheitsfragen zu stellen oder diesen mit einem körperlichen Belang aufzusuchen. Hintergrund dieser Tendenz ist oftmals die Unsicherheit über die Anwesenheit und Autorität des Arztes, die auf manche Patienten einschüchternd wirken und Angst vor einer wertenden Beurteilung hervorrufen kann. Ebenso motiviert die Überlegung, dass das Anliegen nicht wichtig genug erscheinen mag, um einen Arzt aufzusuchen, zur Nutzung eines Chatbots. Die aktuelle Covid-19-Pandemie bekräftigt, dass viele Patienten keinen Hausarzt haben und sich deshalb nicht wohl dabei fühlen, einem Fremden, dem neuen Hausarzt, von ihren Beschwerden zu erzählen.

Auch die Kompetenz eines Chatbots mag eine Motivation darstellen. Gelegentlich sind Patienten der Auffassung, dass Ärzte ihren beruflichen Anforderungen teilweise nicht gerecht werden, ein Chatbot aber jederzeit zuverlässig gute Arbeit leisten könne. Diese Annahme beruht im Wesentlichen auf dem sogenannten 'menschlichen Versagen', weshalb das Vertrauen in eine Maschine als größer eingeschätzt wird. Des Weiteren kann die Motivation auch darin liegen, sich medizinische Zusammenhänge in verständlicher Sprache erklären zu lassen. Informationen können durch den Chatbot ausführlich dargestellt und erklärt werden, anders als in vielen Arztpraxen, in denen ein Problem oft nur knapp bzw. in Fachsprache geschildert wird. Voraussetzung dafür wäre nach Einschätzung der Experten, dass

der Chatbot über ein Gedächtnis verfügt, damit entsprechende Fragen explizit verständlich beantwortet werden können.

Der Anreiz, Informationen durch einen Chatbot zu erhalten, basiert auch auf der unzuverlässigen und nicht vertrauenswürdigen Quelle 'Dr. Google'. Ein Chatbot könnte als seriöse Informationsquelle zum Thema Gesundheit wahrgenommen werden und als 'Tor zum Arztgespräch' dienen, indem er den Patienten zum Beispiel an einen geeigneten Arzt in der Nähe verweist. Vertrauenswürdige Beratung und Hilfestellung, um eine Handlungsempfehlung für den Umgang mit gesundheitlichen Beschwerden zu erhalten, wird somit als eine zentrale Motivation von Patienten gesehen. Dies trifft auch insbesondere bei Auslandsaufenthalten des Patienten zu, während denen mglw. ein deutschsprachiger und vertrauenswürdiger Arzt nicht einfach zu finden ist, der Chatbot jedoch vergleichsweise einfach zu erreichen wäre.

Möglicherweise mag auch eine Motivation für Patienten zur Chatbotnutzung darin liegen, sich zu einem früheren Zeitpunkt mit dem eigenen Gesundheitszustand auseinanderzusetzen, ohne dass bereits ein akutes gesundheitliches Problem vorliegt. Somit könnte der Patient auf Basis der bereitgestellten Informationen eine bessere und einfachere Möglichkeit haben, frühzeitig mehr Verantwortung für die eigene Gesundheit zu übernehmen. Im Idealfall könnten so durch eine frühzeitige Kontaktaufnahme gesundheitliche Konsequenzen vermieden werden.

Zuletzt könnte auch schlicht das Interesse an technischen Neuerungen eine Motivation sein, einen KI-basierten Chatbot auszuprobieren.

4.3 Ergebnisse zu den Erwartungen der Patienten an KI-basierte Chatbots

Insgesamt wurden durch die Experten der Studie 94 potentielle Erwartungen von Patienten an KI-basierte Chatbots genannt, die sich in sieben Kategorien gruppieren lassen.

Zunächst wurde genannt, dass Patienten Bestätigung für das eigene Handeln suchen. Patienten wünschen sich, mit Hilfe der Technologie sicher durch das komplexe Gesundheitswesen geführt zu werden, um das eigene Handeln zu bestätigen. Der Chatbot soll den Patienten somit Orientierung und Unterstützung bieten.

Wesentlich aus Sicht der Experten ist auch die Gebrauchstauglichkeit des Chatbots. Funktionalität, Qualität der Antworten, maschinelle Intelligenz und Benutzerfreundlichkeit stellen dabei zentrale Anforderungen dar. Beispielsweise sollte vermieden werden, dass das Problem mehrfach vom Patienten geschildert oder bei jeder Benutzung die gesamte Krankheitsgeschichte erneut erzählt werden muss. Von einem Chatbot, der auf KI basiert, wird auch erwartet, dass die Redewendung der klassischen Telefonroboter „Ich habe sie nicht verstanden, wiederholen Sie Ihr Anliegen“, keine Verwendung findet, sondern dass stets auf die gestellte Frage geantwortet werden kann, d. h., dass der Chatbot ein auf den einzelnen Patienten bezogenes Gedächtnis besitzt. Dies beinhaltet auch die Erwartung, dass

die vom Chatbot gesammelten Daten auf Wunsch mit dem Arzt geteilt werden können. Der Chatbot sollte auch Auskunft darüber geben können, ob der Patient bei dem von ihm geschilderten Anliegen einen Arzt aufsuchen sollte. Nach einem Arztbesuch besteht umgekehrt die Erwartung, dass der Chatbot die mit dem Arzt besprochenen Themen kennt, um dem Patienten darauf aufbauend weiterhin eine Hilfestellung zu sein. Hinsichtlich der Art der Kommunikation besteht die Erwartung, dass der Chatbot bei einer Problematik sachlich weiterhilft und nicht am Thema vorbei chattet, dass die Antworten des Chatbots sofort kommen, sehr spezifisch sind und gezielt der Fragestellung entsprechen und dabei nicht nur fachlich einwandfrei sind, sondern auch für jedermann verständlich ausgedrückt werden.

Bei der Benutzerfreundlichkeit spielt die Schnelligkeit in der Handhabung eine tragende Rolle. Ein Chatbot sollte leicht zugänglich sein, vorzugsweise auf dem Kanal und dem Endgerät, auf dem sich der Patient zum Zeitpunkt der Suche befindet. Dazu sollte der Chatbot nicht erst gesucht werden müssen, sondern sollte dort platziert werden, wo er benötigt wird, zum Beispiel integriert auf der Website der Arztpraxis. Ferner sollte er einfach, intuitiv und praktisch zu bedienen sein. Dies erfordert die Verfügbarkeit des Chatbots in unterschiedlichen Formen, z. B. als App auf mobilen Endgeräten genauso wie integriert in die Webpage von Gesundheitseinrichtungen. Denkbar wäre auch die Integration in ein sprachgesteuertes System wie z. B. Siri oder Alexa.

Eine weitere wesentliche Erwartung, die von den Experten geäußert wurde, betrifft das Nutzungserlebnis der Patienten. Dieses umfasst deren Wahrnehmungen und Reaktionen bei der Chatbotnutzung sowie das Erleben, welches sich aus der erwarteten Nutzung des Chatbots durch die Patienten ergibt. Dies sind alle Emotionen, Ideen, Vorlieben, Wahrnehmungen, physiologische und psychologische Reaktionen und Verhaltensweisen, die vor, während und nach der Nutzung auftreten (Law et al. 2009: 724). Das Nutzungserlebnis dürfte von großer Wichtigkeit für die Entscheidung, ob der Chatbot erneut genutzt wird, sein. Insbesondere das User Interface, also die Benutzeroberfläche sollte nicht rein technisch, sondern so gestaltet sein, dass sich die Interaktion mit dem Chatbot wie mit einem Menschen anfühlt. Generell wird ein attraktives und modernes Design erwartet. Die Struktur der Benutzeroberfläche sollte eine einfache und intuitive Handhabung ermöglichen.

Bezüglich der Frage, ob der Patient möchte, dass eine Maschine ihm Empathie entgegenbringt, gehen die Meinungen der befragten Experten auseinander. Eine gewisse Vermenschlichung werde wohl gewünscht, im Idealfall soll der Patient gar nicht merken, dass es sich um einen Roboter handelt und der Chatbot Emotionen zeigen sowie die Gefühlslage des Patienten interpretieren und widerspiegeln können. Auf der anderen Seite wurde der Vorbehalt geäußert, dass der Chatbot nicht emphatisch reagieren solle, da Emotionen bei einer Maschine nicht der Realität entsprechen können, sodass dem Patienten letztlich nur etwas vorgespielt werde. Vielmehr solle der Chatbot ausschließlich Fakten erläutern. Ein Lösungsansatz könnte sein, dass der Patient selbst einstellen können sollte, wie sich der Chatbot verhalten soll. Einig sind sich die Experten darin, dass der Chatbot sich bei der erstmaligen Nutzung

klar als solcher zu erkennen geben soll, deutlich machen soll, dass die Maschine auf Basis eines Algorithmus agiert und dass der Patient nicht in dem Glauben gelassen werden darf, dass er mit ärztlichem Personal spricht.

Als sehr wesentlich wird durch die Experten auch gesehen, dass Datensicherheit und Datenschutz bei der Chatbotnutzung unbedingt zu gewährleisten sind. Eine Option wäre, dass der Chatbot mit einem Zertifikat ausgezeichnet wird, das garantiert, dass Daten sicher verwahrt werden und einfach nachzuverfolgen sind. Datenschutz kann in der Wahrnehmung der Patienten am besten erreicht werden, wenn es sich um ein deutsches Unternehmen handelt, das den Chatbot betreibt, sodass die Daten auch nach deutschem Recht verwaltet werden. Speziell wird die Erwartung vermutet, dass ohne ausdrückliche Zustimmung Versicherungen keinen Einblick in die Gesundheitsdaten des Patienten erhalten dürfen. In diesem Zusammenhang wird Transparenz bei der Chatbotnutzung erwartet, um die Vertrauenswürdigkeit des Chatbots zu gewährleisten. Ausreichende Informationen von Seiten des Chatbotbetreibers sollten daher bereitgestellt werden. Es wird auch erwartet, dass der Chatbot seine funktionalen Grenzen klar kennt und kommuniziert. Für den Fall, dass die Technik des Chatbots ausfällt, wird eine Hotline für die Patienten gewünscht.

4.4 Ergebnisse zu den Hemmnissen der Chatbotnutzung

Die Nutzung digitaler Technologien wird generell durch die Experten als altersgruppenabhängig gesehen. Ein hohes Alter stellt dabei tendenziell ein Hemmnis für die Chatbotnutzung dar, dies ginge soweit, dass die ältere Generation faktisch von der Nutzung ausgeschlossen würde. Dies liege weitgehend an der Zugänglichkeit, die bei mangelndem Technik-Know-how nicht gegeben sei. Grundsätzlich gilt, dass diejenigen Personen, die nicht über die notwendige Technik wie z. B. Smartphones, Tablets oder Laptops verfügen oder mit deren Umgang nicht ausreichend versiert sind, von der Chatbotnutzung ausgeschlossen sind.

Auch der Aufwand, um den Chatbot nutzen zu können, kann ein Hemmnis darstellen. Eine langwierige sowie komplexe Suche, um überhaupt mit dem Chatbot kommunizieren zu können, schließt die Nutzung weitgehend aus.

Ein weiteres Hindernis ist aus Sicht der Experten der Mangel an persönlichem Kontakt. Von einer Maschine fühlten sich viele Patienten nicht gleichermaßen angesprochen wie von einem Arzt, der in einem persönlichen Gespräch Emotionen und Empathie vermitteln kann. Patienten möchten trotz vorheriger Kommunikation mit einem Chatbot den gesundheitsbezogenen Sachverhalt letztlich doch nochmal mit einem Arzt besprechen.

Mangelndes Vertrauen in das System stellt ein weiteres potentiell Hemmnis dar. Werbeanzeigen vor oder während der Chatbotnutzung können dem Patienten implizieren, dass der Chatbot nicht aus gesundheitlichen Motiven, sondern aus kommerziellen Gründen existiert. Das User Interface spielt dabei

eine wesentliche Rolle, ob der Chatbot in Anspruch genommen wird und ob das System und das Unternehmen, das hinter dem Chatbot steht, vertrauenswürdig erscheinen.

Für einen Großteil der Patienten spielt aus Sicht der Experten der Datenschutz eine entscheidende Rolle und stellt daher auch ein Hemmnis für die Nutzung dar, wenn ihre diesbezüglichen Erwartungen nicht erfüllt wird. Hier stellt sich insbesondere die Frage, wo die Daten gespeichert werden und wie mit ihnen umgegangen wird. Auf den Schutz der Daten vor einem Zugriff durch Versicherungen wurde oben bereits verwiesen.

4.5 Weitere Ergebnisse der Studie

Neben Antworten auf die vier Kernfragen der Studie lieferten die Experten auch weitere Gedanken und Ansatzpunkte zur Anwendung von KI-basierten Chatbots im Gesundheitswesen. Der zentrale Gedanke dabei ist, dass Chatbots als seriöse, zuverlässige und vertrauenswürdige Informationsquelle Alternative zu Google und anderen Internetquellen dienen kann. Das Haupteinsatzfeld von Chatbots ist der Patientenerstkontakt. Dabei kann der Chatbot die Anamnese, Diagnostik und (Selbst-)Diagnose allein oder in Kooperation mit einem Arzt durchführen. Chatbots können zur Prävention eingesetzt werden, wenn sich Patienten über gesundheitsfördernde Maßnahmen informieren möchten. So kann der Patient mehr Verantwortung für seine Gesundheit übernehmen. Der Chatbot kann dazu weitergehende Informationen zu Ernährung, Bewegung oder Schlaf zur Verfügung stellen sowie den Patienten anleiten, seinen Gesundheitszustand zu optimieren, indem er über Konsequenzen von seinem Verhalten informiert und mit Hilfe von Anleitungen zu Verhaltensänderungen motiviert wird.

Eine weitere mögliche Anwendung ist die Unterstützung bei chronischen Krankheiten wie Diabetes, beginnender Demenz oder Langzeiterkrankungen wie Krebs. Die Betroffenen sind dann ohnehin schon im Austausch mit Ärzten und könnten den Chatbot als weitere Hilfe in Anspruch nehmen. Der Chatbot könnte den Erkrankten über die Krankheit aufklären und ihn an die neuen Lebensumstände heranzuführen oder ihn an die Einnahme von Medikamenten erinnern. Die mit dem Chatbot ausgetauschten Informationen können an den Arzt weitergeleitet werden, damit dieser den Gesundheitszustand besser im Auge behalten kann. Auch im Bereich der Bilderkennung lässt sich der Chatbot sinnvoll einsetzen, Algorithmen können aufgrund der gesammelten Daten bei der Bilderkennung leistungsfähiger sein als der Arzt selbst.

Ein weiterer denkbarer Einsatzbereich von Chatbots sind organisatorische Tätigkeiten etwa in der Notaufnahme oder im Wartezimmer. Chatbots könnten eingesetzt werden zur Durchführung der Anamnese, die dem Arzt so vorab vorliegt. Dies vereinfacht die Folgeprozesse im Praxisalltag und kann zu einer effizienteren Aufnahme und Sprechstunde beitragen.

Der Chatbot könnte jedoch nicht nur auf Patientenseite, sondern auch als Unterstützung der Ärzte eingesetzt werden, da er aufgrund von großen Datenmengen über einen enormen Erfahrungsschatz

verfügt. Es besteht die Möglichkeit, dass sich der Arzt nach dem Gespräch mit dem Patienten eine zweite Meinung durch den Chatbot einholt, um die Diagnose zu überprüfen, eine genauere Diagnose zu treffen oder alternative Therapieansätze zu finden.

Die an dieser Arbeit beteiligten Experten äußerten sich auch zu den Grenzen des Einsatzes von KI-basierten Chatbots. Zunächst ist hier der gesetzliche Rahmen der Datenschutzbestimmungen zu nennen. Der Verarbeitung von personenbezogenen Daten sind in Deutschland enge Grenzen gesetzt. Eine weitere Grenze für Chatbots stellt das Risiko von Fehlinterpretationen und Fehlinformationen dar. Es besteht die Gefahr einer Fehldiagnose, da der Chatbot nicht über die menschlichen Sinne des Sehens, Riechens, Fühlens und Hörens verfügt. Bei einer Diagnose spielen diese Sinne jedoch eine entscheidende Rolle, der Chatbot könnte entsprechend die vom Patienten gegebenen Informationen falsch interpretieren. Bedeutsam ist auch der Umstand, dass Menschen unterschiedliche Erwartungen, Wahrnehmungen und ein unterschiedliches Körpergefühl haben. Chatbots sind nur sehr begrenzt in der Lage, verschiedene Typen von Patienten zu berücksichtigen, die Selbsteinschätzungen von Patienten kritisch zu bewerten und zu differenzieren, um einen objektiven Befund abzuleiten.

Es wurde auch die Frage aufgeworfen, ob Chatbots den strukturierten und gewohnten Tagesablauf der Krankenhaus- bzw. Arztpraxis stören, da die Integration in den regulären Tagesablauf nicht ohne weiteres möglich ist, ob also Chatbots wirklich eine Kostenersparnis bedeuten. In diesem Zusammenhang ist auch fraglich, ob sich durch Chatbots auch die Art der Kommunikation und Zusammenarbeit zwischen Arzt und Patient nachhaltig ändern wird.

5 Fazit & Ausblick

Trotz der hohen Erwartungen hat ein KI-basierter Chatbot nach aktuellem technischem Stand nur begrenzte Möglichkeiten, sodass der Chatbot den Arztbesuch letztlich nicht ersetzen kann und soll. Umgekehrt gibt es trotz der möglichen Probleme ein großes Potenzial für die Nutzung solcher Chatbots im Gesundheitswesen. Die Zukunft von Chatbots ist grundsätzlich vielversprechend, da die Durchdringungsrate in der Gesellschaft von neuen digitalen Technologien stetig steigt. Durch das Sammeln großer Mengen an Daten und Diagnosen ist ein Chatbot in der Lage, in kürzester Zeit einen größeren Erfahrungsschatz aufzubauen als ein einzelner Arzt. Umfangreichere Analysen und Erfahrungsberichte über Krankheiten können erstellt werden, um kostenoptimierte und effizientere Therapiemöglichkeiten anzubieten. Grundsätzlich sind Chatbots auch als Instrument zur Förderung des Gesundheitsbewusstseins der Bevölkerung geeignet.

Von wesentlicher Bedeutung, um diese Potentiale nutzen zu können, wird sein, die Nutzungshemmnisse zu überwinden. Insbesondere die Verfügbarkeit und Nutzungskompetenz digitaler Kommunikationskanäle erscheint hier wesentlich. Eine sich stetig verbessernde medizinische Kompetenz, d. h. zunehmende Intelligenz in der Verarbeitung auch komplexer Anamnesen, eine enge Interaktion mit dem

behandelnden Arzt, hohe Anwendungsfreundlichkeit, verlässliche Funktionalität sowie Sicherheit der Datenspeicherung und -verarbeitung verbunden mit patientengerechter Sprache und Elementen menschlicher Empathie weisen den Weg zu steigendem Vertrauen der Patienten in KI-basierte Chatbots.

Chatbots sind am besten zur Unterstützung von Ärzten eingesetzt, anstatt sie zu ersetzen. Es scheint nicht die Gefahr zu bestehen, dass Ärzte durch Chatbots aus dem Gesundheitssystem verdrängt werden, vielmehr besteht der Wunsch, Gesundheit und Technologie zu verbinden, sodass langfristig die Nutzung von KI-basierten Chatbots in eine optimierte Diagnose und Therapie integriert wird.

Die Ergebnisse dieser Studie basieren auf einer explorativen Erhebung unter acht Experten. Naturgemäß konnten so nur subjektive Ansatzpunkte aus deren Blickwinkel gewonnen werden. In einem nächsten Schritt wäre es sinnvoll, im Zuge einer qualitativen Studie mit einer deutlich größeren Stichprobe aus Patienten, Vertretern des Gesundheitswesens und Entwicklern von KI-basierten Chatbots die Ergebnisse dieser Studie zu verifizieren, Erkenntnisse über den Grad der Bedeutung einzelner Aspekte zu gewinnen und hieraus Ableitungen zu treffen, welche Faktoren besonders erfolgsrelevant sind und somit klare Handlungsempfehlungen für die Entwicklung und Nutzung von KI-basierten Chatbots zu gewinnen.

Literaturverzeichnis

- Abashev, A.; Grigoryev, R.; Grigorian, K.; Boyko, V. (2017): Programming Tools for Messenger-Based Chatbot System Organization: Implication for Outpatient and Translational Medicines. In: *BioNanoScience* (2017), Nr. 7, S. 403–407.
- Bendig, E.; Erb, B.; Schulze-Thuesing, L.; Baumeister, H. (2019): Die nächste Generation: Chatbots in der klinischen Psychologie und Psychotherapie zur Förderung mentaler Gesundheit – Ein Scoping-Review. In: *Verhaltenstherapie* (2019), Nr. 29, S. 266–280.
- Chaix, B.; Bibault, J.-E.; Pienkowski, A.; Delamon, G.; Guillemassé, A.; Nectoux, P.; Brouard, B. (2019): When Chatbots Meet Patients: One-Year Prospective Study of Conversations Between Patients With Breast Cancer and a Chatbot. In: *JMIR Cancer* (2019), Nr. 5.
- Crutzen, R.; Peters, G.-J. Y.; Portugal, S. D.; Fisser, E. M.; Grolleman, J. J. (2011): An Artificially Intelligent Chat Agent that Answers Adolescents' Questions Related to Sex, Drugs, and Alcohol: An Exploratory Study. In: *Journal of Adolescent Health* (2011), Nr. 48, S. 514–519.
- Dubey, S. K.; Rana, A. (2010): Analytical Roadmap to Usability Definitions and Decompositions. In: *International Journal of Engineering Science and Technology* (2010), Nr. 2.
- Dudenredaktion (Bibliographisches Institut) (Hrsg.) (2015): *Duden deutsches Universalwörterbuch*. 8., überarb. und erw. Aufl., Dudenverlag.
- Espinoza, J.; Crown, K.; Kulkarni, O. (2020): A Guide to Chatbots for COVID-19 Screening at Pediatric Health Care Facilities. In: *JMIR Public Health and Surveillance* (2020), Nr. 6(2).

- Fitzpatrick, K. K.; Darcy, A.; Vierhile, M. (2017): Delivering Cognitive Behavior Therapy to Young Adults With Symptoms of Depression and Anxiety Using a Fully Automated Conversational Agent (Woebot): A Randomized Controlled Trial. In: *JMIR Mental Health* (2017), Nr. 4.
- Holmes, S.; Moorhead, A.; Bond, R.; Zheng, H.; Coates, V.; Mctear, M. (2019): Usability testing of a healthcare chatbot: Can we use conventional methods to assess conversational user interfaces? In: *Proceedings of the 31st European Conference on Cognitive Ergonomics* (2019), S. 207–214.
- Laranjo, L.; Dunn, A. G.; Tong, H. L.; Kocaballi, A. B.; Chen, J.; Bashir, R.; Surian, D.; Gallego, B.; Magrabi, F.; Lau, A. Y. S.; Coiera, E. (2018): Conversational agents in healthcare: a systematic review. In: *Journal of the American Medical Informatics Association* (2018), Nr. 25, S. 1248–1258.
- Laumer, S.; Maier, C.; Gubler, F. T. (2019): Chatbot acceptance in Healthcare: Explaining user adoption of conversational Agents for Disease Diagnosis (2019).
- Law, E. L.-C.; Roto, V.; Hassenzahl, M.; Vermeeren, A. P. O. S.; Kort, J. (2009): Understanding, scoping and defining user experience: a survey approach. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (2009), S. 719–728.
- Liebrich, F. (2017): *Digitale Medienprodukte in der Arzt-Patienten-Kommunikation*. Wiesbaden: Springer Fachmedien.
- Ly, K. H.; Ly, A.-M.; Andersson, G. (2017): A fully automated conversational agent for promoting 46 mental well-being: A pilot RCT using mixed methods. In: *Internet Interventions* (2017), Nr. 10, S. 39–46.
- Mayring, P. (2015): *Qualitative Inhaltsanalyse: Grundlagen und Techniken*. 12., überarb. Aufl., Weinheim: Beltz.
- Meier, P.; Beinke, J. H.; Fitte, C.; Behne, A.; Teuteberg, F. (2019): FeelFit – Design and Evaluation of a Conversational Agent to Enhance Health Awareness. In: *FeelFit – Conversational Agent to Enhance Health Awareness* (2019).
- Nadarzynski, T.; Miles, O.; Cowie, A.; Ridge, D. (2019): Acceptability of artificial intelligence (AI)-led chatbot services in healthcare: A mixed-methods study. In: *DIGITAL HEALTH* (2019), Nr. 5.
- Palanica, A.; Flaschner, P.; Thommandram, A.; Li, M.; Fossat, Y. (2019): Physicians' Perceptions of Chatbots in Health Care: Cross-Sectional Web-Based Survey. In: *Journal of Medical Internet Research* (2019), Nr. 21.
- Pravettoni, G.; Triberti, S. (2020): *P5 eHealth: An Agenda for the Health Technologies of the Future*. Springer International Publishing.
- Richter, L.; Silberzahn, T. (2020): Germany's e-health infrastructure strengthens, but digital uptake is lagging. (2020).
- Seebach, N.; Wasilewski, L. (2021): *Digitaler Puls: warum der Gesundheitsmarkt jetzt digital handeln muss!* Bern: Hogrefe.
- ter Stal, S.; Kramer, L. L.; Tabak, M.; op den Akker, H.; Hermens, H. (2020): Design Features of Embodied Conversational Agents in eHealth: a Literature Review. In: *International Journal of Human-Computer Studies* (2020), Nr. 138.

Tudor Car, L.; Dhinakaran, D. A.; Kyaw, B. M.; Kowatsch, T.; Joty, S.; Theng, Y.-L.; Atun, R. (2020): Conversational Agents in Health Care: Scoping Review and Conceptual Analysis. In: Journal of Medical Internet Research (2020), Nr. 22.

Zhang, J.; Oh, Y. J.; Lange, P.; Yu, Z.; Fukuoka, Y. (2020): Artificial Intelligence Chatbot Behavior Change Model for Designing Artificial Intelligence Chatbots to Promote Physical Activity and a Healthy Diet: Viewpoint. In: Journal of Medical Internet Research (2020), Nr. 22.

Autoren

Luisa **de Alzaga Achter** ist Junior Operations Managerin bei der Climedo Health GmbH, einem Health-Tech-Start-up in München. Während ihres Bachelorstudiums in International Management an der International School of Management (ISM) sammelte sie Berufserfahrung in verschiedenen Branchen und Bereichen. Fasziniert durch die Möglichkeiten der Digitalisierung im Gesundheitswesens schrieb sie ihre Bachelorarbeit über die Patientenerwartungen an KI-basierte Chatbots. Parallel dazu hat sie den Verein Munich Roots in Education mit gegründet. Als Vizepräsidentin baut sie den Förderverein weiter aus mit der Vision, die Chancengleichheit im deutsche Bildungssystem zu verbessern.



Daniel-Ulf **Khafif** ist Leiter für Business Development und Spin-Offs an der Fakultät für Elektrotechnik, IT und Nachrichtentechnik an der Technischen Universität Berlin (TUB) und Dozent für den Studiengang Strategisches Marketing-Management an der International School of Management (ISM) in München. In seinem Magisterstudium der Komparatistik, Romanistik und Kunstgeschichte entwickelte Daniel Khafif seinen interdisziplinären Arbeitsschwerpunkt zur kulturellen und ökonomischen Wirkung technischer Innovation. Der Journalist (DJV), Buchautor (KUNST BAU PUNK) und Manager für Marketing und Kommunikation (Goldenen IF Award Industriekommunikation 2008) führt seit 25 Jahren für Unternehmen der Mobility, Luft- und Raumfahrt, Logistik und IT die Schnittstellenkommunikation zwischen technischer Innovation, Produktstrategie und Marktintegration.



Prof. Dr. Klaus **Mühlbäck** ist Leiter des Master of Arts Studiengangs Strategisches Marketing Management (English Trail) an der International School of Management (ISM) in München. Derzeit ist er Mitglied des Senats der Hochschule. Klaus Mühlbäck hat einen Abschluss in Marketing Management und promovierte im Bereich des internationalen Handels mit dem Schwerpunkt Internationalisierungsstrategien mittelständischer Industrieunternehmen. Mit mehr als 20 Jahren Berufserfahrung in einem hochgradig interkulturellen Umfeld liegen seine Lehr- und Forschungsschwerpunkte in den Bereichen strategisches Marketingmanagement, interkulturelles Marketing und Konsumentenverhalten.



Schuck, Katharina A.; Perret, Jens K.

Optimizing Production of Fashion Goods as Means to a More Sustainable Garment Industry – An Assessment of the Literature

Abstract

Objectives:

The overall objective is to create awareness that production scheduling methods represent highly valuable tools in contributing to a more sustainable garment manufacturing. Further, a structured illustration of existing objectives of production scheduling is provided by grouping them according to their influence on the production runs. Based on this, it is demonstrated how each of the six groups would entail beneficial sustainability effects.

Theoretical Framework:

The theoretical framework is based on six decades of research, which produced a broad range of contributions to production scheduling for different production layouts and designs (flow-shop, job-shop, assembly lines), covering the full range of fashion production designs.

Method:

Methodologically, a literature review and an accompanied analysis are provided.

Results/Implications:

While the economic advantages of production planning are obvious, the discussion revealed that literature already contains considerable potentials for optimizing garment production, which, however, need to be applied to the fashion context, especially when it comes to enhanced sustainable business practices. As the link between optimization and holistic sustainability has received limited attention, this article presents a comprehensive conceptual view on the topic and provides decision makers and researchers with valuable inputs on which optimization approaches can support a company's sustainability efforts.

Keywords:

Sustainability, fashion, optimization, job-shop, flow-shop, assembly line

1 Introduction

The fashion market is a highly competitive, globally connected, diverse, and volatile market. It is estimated that around 40 million people are working in the fashion industry worldwide, with a large proportion probably remaining unaccounted for due to the complexity of the supply chain and a lack of international control mechanisms. Approximately 43 % of sales are generated with so-called 'entry-price' products, with large and internationally operating 'Giant Players' having the upper hand (BCG 2019). In the past, it was especially this low-priced and fast-moving part of the industry – also called 'fast fashion' which includes globally popular clothing brands such as Zara and H&M – that faced criticism due to a number of issues regarding environmental pollution and the socially questionable treatment of workers. According to the UN, the fashion industry is responsible for 20 % of water pollution, making it the second-largest water polluter and thus a very influential player (UNEP 2022).

Although being part of the fashion industry is extremely lucrative for numerous participants, the COVID-19 pandemic in particular revealed the industry's fragility to outside economic shocks. As consumers demanded less clothing, some clothing companies canceled orders and refused to pay for already ordered goods, while others tried to find new business fields and activities (Sumarliah et al. 2021). As the fashion industry is characterized by the fact that a number of supply chain partners – primarily the ones in the Southeast-Asian production locations – are dependent on the apparel companies, the situation was quite challenging for many of them (Brydges et al. 2020). Hence, the pandemic emphasized that the production of many fashion items is socially problematic (Kulsum 2020). At the same time, practitioners, researchers, and branch experts also brought the environmental issue of fashion production once again into the spotlight, catalyzed by consumers demanding security and durability in pandemic times, not only with regard to their daily lives, but also regarding their consumption (Kim 2021). Initially, the manufacturing process of garments was understood as a creative process characterized by craftsmanship and skill. Today, for many high-street fashion companies, it is especially the economic motive that is of primary concern (BoF & McKinsey 2021), and, only gradually, is the awareness that sustainable fashion can be economically advantageous while – at the same time – environmentally and socially friendly, taking ground in manufacturers' minds.

The economic motive is the one primarily targeted by applications of operations research in production planning, but economic goals do not necessarily need to be opposed to sustainability goals in general (Alberti & Varon Garrido 2017). Thus, a synthesis of both might be a solution for the apparel industry. Producing more efficiently can increase their profit margin while at the same time making the company more sustainable. Aside from operations research a number of other approaches exist to target the same problems, but as long as the relevant sources of information are certain to a degree mathematical optimization will provide the superior results.

Considering the supply chain of wearing apparel in full then the final production only accounts for a limited share and thus sustainability issues of fashion production can not be solved by optimization of the final production alone. Nevertheless, optimization of the final production is the part of the supply chain under the direct influence of the producing company itself and therefore it is the ideal starting point for sustainability-oriented changes in fashion production as such.

Taking a look at previous findings on the topic of garment manufacturing and optimization, there are different angles being taken. For some researchers, the place of manufacture as well as the materials used are of concern: Examples include the extent to which the production of clothing is compatible with the principles of social and environmental sustainability, highlighting the perspective of workers (Arvidsson et al. 2010; Perry & Towers 2013) as well as the resources and yarns used (Debnath 2016; Gwilt 2020; Niinimäki et al. 2020). While describing which business models could be conceivable for improving sustainable business activity, Pal and Gander (2018) provide one of the few contributions discussing environmentally sustainable fashion production from an economic perspective. Other studies are more concerned with the whole supply chain management: On a conceptual level, Bruce et al. (2004) address the issue of just-in-time delivery and argue that this might lead to achieving agility and flexibility in production, which – according to many authors – is mandatory for successful fashion production (Barnes & Lea-Greenwood 2006; de Oliveira Neto et al. 2019; Perry 2017). Further, Barnes and Lea-Greenwood (2006) emphasize the advantages of a vertical business model when it comes to enabling the most efficient supply chain management (and subsequently manufacturing). Ways to improve the actual manufacturing process are also provided: Liu et al. (2020) focused on the mass customization production format and delivered ideas that can contribute to efficiency improvement. Perret et al. (2022) concentrated on the same production format and delivered a cost-based optimization approach based on the concept of level scheduling. Further, Fani (2018) provides a scheduling approach similar to the one by Perret et al. (2022) but with a different array of objectives and not based on the level scheduling paradigm. Eira et al. (2015) conducted a case study of a clothing manufacturer and provided tailored waste elimination opportunities. Ait-Alla et al. (2014) present a mathematical model for fashion apparel suppliers to support them in achieving improvement regarding the allocation of production orders, whereas Choi et al. (2014) provide demand-focused forecasting tools for the fashion industry, i.e. mainly the manufacturer's side.

However, scientific studies dedicated to the optimization possibilities of garment manufacturers are limited in scope and often outdated regarding their methodology. It is the fact that fashion manufacturers changed little – despite years of drawing attention to the (social and environmental) problems in their production locations – shows that there is still room for improvement by integrating sustainability considerations into the discussion of the use of operations research in apparel manufacturing. Perret (2023) is among the few studies that take this aspect into consideration.

A look at other industries reveals that, in contrast to a plentiful breadth of literature, the set of objectives implemented in any of the major production layouts – job-shop, flow-shop and assembly line – is limited. While very few publications focus on applications of operations research in the fashion industry (Fani 2018; Perret et al. 2022), the existing studies and their accompanied methodologies, conducted for other industries, apply to the fashion industry nonetheless. The research objective of this paper therefore lies in delivering a summary of implemented objectives in job-shop, flow-shop and assembly line scheduling across literature and grouping them into sets with the same overarching goal. Each of the groups is discussed, not only in terms of its underlying objective, but also in respect of the benefits it offers to producers, i.e. particularly regarding social and ecological sustainability issues. Even though this study does not deliver a distinct optimization approach in itself, it provides a meta-study of objectives, offering both researchers and practitioners a toolbox to construct their own multi-objective optimization problems tailored to their particular needs and constraints. By referring to different production layouts, it presents easily applicable improvement options that not only lead to an economically more lucrative production process but also to a more environmentally and socially sustainable one. Hence, the problem as such might be well-known among practitioners in the fashion industry, it has not yet been considered in the academic literature. Thus, the present article contributes to the limited scientific literature base of fashion production planning and sequencing while incorporating practical relevance regarding sustainability issues.

2 Background on Garment Production

2.1 From an Industry Perspective

The manufacturing process of garments can happen in different formats, which require different production layouts. Overall, the production format chosen also has to do with the desired pricing of the product, as different formats demand different budgets, which is mainly due to the quantity that can be produced (Bevilacqua et al. 2013; Cachon & Swinney 2011). One of these possibilities is ‘custom tailoring’. It is largely carried out in a job-shop scheduling in which individual work tasks are done after one another without many processes being clearly defined in advance. In custom tailoring, manual work is also frequently used, as there is hardly any possibility of standardizing the processes due to the customized products and the individually designed tailor stores (Parveen & Ullah 2010). Hence, there are often qualified seamstresses working in this production format. The advantage of this method is that there is a high level of customer integration and, thus, a high probability of satisfaction and loyalty on the consumer side. In terms of production, the lack of standardization and the low production frequency result in considerable time and financial requirements, which are consequently also passed on to the products’ retail price. For this reason, tailored products tend to be in the premium or luxury price segment.

A very large proportion of companies focusing on low-priced and trend-oriented fashion items produce in 'mass productions'. Mass production is traditionally organized as a 'push-system'. The focus here is on a precise forecast, as the products and production quantities are planned on the basis of previous seasons' sales (Liu et al. 2020; Yeung et al. 2010). The production sites in this format are often located in low-wage countries, as most steps of the actual garment manufacture are carried out by workers on individual sewing machines (de Oliveira Neto et al. 2019; Niinimäki et al. 2020). In order to make the production as efficient as possible, there is usually a line of sewers strung together, all of whom have precisely defined jobs. By doing that, a highly beneficial standardized process is possible – beneficial as standardization favours speed. A possible scenario is that one sewer could only be responsible for the right arms of a shirt, whereas the next one in line is only putting the finished arms and middle parts together. Further, efficient inventory management is crucial for mass-producing fashion companies. Products manufactured in distant production facilities such as India, China, or Bangladesh are produced with a long lead time to their actual time of sale because otherwise, there would be problems due to the long transport distances (Camargo et al. 2020; Christopher et al. 2004). While this format has the advantage of standardized operations, it is error-prone and not very customer-centric: If a trend changes quickly (which is very often the case in high-street fashion) or a product does not arrive as planned, it accompanies the backlash that large quantities of goods remain unused and have to be destroyed, offered at large discounts, or be discarded (DW 2021). Alongside the spontaneously changing orders, there is enormous pressure on both the manufacturing sites and the sewers because there is only a very narrow timeframe where a trend (and with it, the product) is relevant for the consumers, i.e. for the fashion company (Christopher et al. 2004; Rafi-Ul-Shan et al. 2017).

In addition, there is also a kind of hybrid form of the two production formats described that is called 'mass customization'. In contrast to mass production, mass customization is oriented towards a 'pull-system' which requires a regular exchange of information between the various parties involved in the supply chain (Yeung et al. 2010). To respond to the fast-moving fashion industry, more and more companies have recently decided to produce in this specific format (Xu et al. 2020). As far as the production process is concerned, a final product is assembled only when a customer requests it. The choice of production layout is strongly related to the level of customization, meaning that more individuality and customizability also require more individually adapted production steps. Within this format, fashion producers greatly benefit from fewer customer returns because they influence the products' designs themselves. Since only the exact demand is produced, large overstocks can be eliminated, and storage costs as well as potential waste will be reduced. Nevertheless, this format also has a downside: It requires extreme flexibility and generates operational complexity, which is reflected in both financial expenditure and know-how requirements (Liu et al. 2020; Xu et al. 2020; Yang et al. 2015).

Note that the following study can be read as a literature review of the different objectives used in mathematical production planning and sequencing in the context of the three main production layouts

(Section 3.1). Second, it can be read as a first qualitative clustering of the objectives in use in this field of mathematical optimization (Section 3.2). Alternatively, it can also be read, and this is its main objective, as a discussion of the sustainability potentials inherent in the objectives, which primarily target output- or cost-related aspects (Section 3.3). While parts of this discussion are already known to practitioners or researchers, this study combines the insights into one comprehensive overview.

2.2 From an Operations Research Perspective

Regardless of the assumed objective, production scheduling in the fashion industry would usually take place in one of three possible layouts: flow-shop, job-shop or assembly line (Blazewicz et al. 2019). In all three situations, a number of tasks need to be performed to finish the manufacture of a pre-defined product, including different variants of the same product at this point. These tasks are assigned to different stations or workers involved in the production process, such as sewers in a sewing room. The assignment of tasks to stations is not made randomly but is determined by a given set of precedent constraints in addition to the assumed goals, i.e. certain tasks must be completed before others can be started (Scholl 1999). Without these precedence constraints, all three types could be subsumed as ‘open-shop problems’ (Blazewicz et al. 2019). The three approaches differ regarding the flow of goods through the production process, particularly resulting from whether each product requires all tasks to be performed and whether buffers or inventories are allowed between stations.

In a flow-shop environment, buffers are allowed to exist between stations, and each product is assumed to have the same operations done to it (Blazewicz et al. 2019). In this context, there are two main challenges arising: The first one emanates from the assignment of tasks to the stations (also referred to as balancing), and the second contains the order to process a series of orders at one of the stations at a specific system layout and station design (Scholl 1999).

Even though smart alternatives are possible with an assembly line, it is usually assumed that each product visits every station, even if no tasks are performed at that station. This case, however, would point to poor balancing in the balancing phase of the assembly line. Where flow-shop production allows for buffers or inventories to exist between stations in general, assembly lines (both in practical terms and mathematical theory) are constructed to avoid buffers and assure a continuous flow of goods via a continuously flowing line (Scholl 1999). Consequently, the main objectives of the two production types are quite similar. Job-shop production, in contrast, does not assume that each product has to visit each of the stations and that each task needs to be executed. Ultimately, it allows for the existence of buffers between stations or different stations in the product run (Blazewicz et al. 2019). A special case applies when there are a number of stations that all perform the same tasks, and the product is completed after all the required stations have been visited. In this case, the problem reduces to a scheduling problem on parallel machines Blazewicz et al. (2019). Generally, the particular

challenge in this situation is two-fold: On the one hand, there is the question of which order should be assigned to which station. On the other hand, however, it is also a matter of which order the jobs assigned to the stations are to be processed, if this option is permitted.

With regard to the fashion context, neither of the three approaches covers fully custom-made products like unique designer pieces. This area, however, can be excluded from this discussion without loss, as the creative and artistic dimension supersedes the planning aspect of the production anyway. For luxury fashion, it seems prudent to refer to a job-shop design, especially since job-shop production usually assumes that each station represents a specific area of expertise and cycle times across stations are not necessarily identical. The derived scheduling problem on parallel machines, where the same products are produced on a number of identical parallel stations, is reminiscent of the classical sweat-shop design, particularly used to generate mass-produced fast fashion goods or simple basic goods, like standard white t-shirts. Products at intermediate and affordable levels are usually mass-produced, but the trend is also continuously moving towards customization at these price levels (Liu et al. 2020). While other sectors, such as the automotive industry, have demonstrated that assembly lines are extremely effective in this environment (Scholl 1999), this has also been proven for the fashion segment (Perret et al. 2022).

Comparing classical flow-shop and assembly line problems in an environment where mass customization (pull production) persists, an assembly line-based approach provides better results (Scholl 1999), whereas in the classical mass production environment (push production), flow-shop approaches might fit better. With ever decreasing fashion cycle times (Choi et al. 2014) and the need to produce a new collection in less than a couple of weeks, it seems more and more critical that both objectives – balancing and sequencing – are jointly considered in a comprehensive approach. In the context of assembly lines, Battini et al. (2009), Defersha and Mohebalizadehgashti (2018) and Lopes et al. (2020) are among the recent studies to present such a joint approach. In the context of fashion production planning, this aspect has not yet been considered. However, since flow-shop and assembly line problems are closely linked, approaches that work in one of the two environments can oftentimes be modified to work in the other environment as well.

In summary, all three approaches to production planning find their application in the fashion industry. Therefore, the following section considers all three approaches, accompanied by a review of objectives used in the context of production sequencing. The focus on production sequencing is favored over production balancing since sequencing provides operative insights, whereas balancing provides strategic insights. Even if the trend is moving towards a decrease in fashion cycle times, the operative perspective is the one that offers the most approaches for making the whole production process more sustainable. The most important aspect of sustainability in the context of balancing is to set realistic task and cycle times that do not overburden workers. The determination of realistic times, however,

is not a question of the balancing approach per se but of the preceding data collection and measurements.

3 Optimization Approaches

3.1 Realized Objectives

Over the course of the last six decades, a broad range of articles on the scheduling of job-shops, flow-shops, or assembly lines were published. However, the implemented objectives – oftentimes propagated as a multi-objective version – remain mostly the same. The main objective in job-shop and flow-shop scheduling is minimizing the makespan, whereas in assembly line sequencing, it is balancing the part usage or workloads, i.e. minimizing work overloads. Table 1 summarizes the objectives used in relevant scheduling problems. For flow-shop scheduling, the objectives result from the recent and exhaustive literature reviews by Rossit et al. (2018) and Komaki et al. (2019), whereas the objectives for job-shop scheduling result from the literature reviews by Chaudhry and Khan (2016) and more recently the one by Zhang et al. (2019). Considering that no inherently new objectives have been introduced in assembly line sequencing in recent years, the literature reviews by Dhamala and Kubiak (2005) and Boysen et al. (2009) are still suitable summaries of the state-of-the-art in assembly line sequencing, particularly when considering the implemented objectives. In addition to each objective in table 1, an exemplary article is cited where the objective is implemented, even if it has only been part of an objective function consisting of multiple sub-objectives.

Table 1: *Implemented Objectives Across Literature*

Flow-Shops	Relating Articles
Makespan	Benavides/Ritt (2016)
Total Completion Time	Al-Anzi/Allahverdi (2013)
Lateness	Al-Anzi/Allahverdi (2007)
Tardiness	Ziaee (2013)
Holding Costs	Park/Kim (1999)
Lead Time	Blocher/Chhajed (2008)
Shipping Costs	Kazemi et al. (2017)
Flow Time	Pugazhendhi et al. (2004)
Changeover or Setup Costs	Grau et al. (1995)
Firm Income	Doganis et al. (2005)
Job-Shops	Relating Articles
Makespan	Sun et al. (2010)
Mean Completion Time	Nie et al. (2013)
Production Costs	Lan et al. (2010)
Balance Workload	Yuan/Xu (2013)
Resource Transition Costs	Lin et al. (2012)
Tardiness Costs	Chen et al. (2012)
Lateness Costs	Vilcot/Billaut (2011)
Mileage of Vehicles	Yan et al. (2022)
Machine Utilization Rate	Chen/Chen (2008)
Production Efficiency	Fattahi/Fallahi (2010)
Mean Flow Time	Nie et al. (2013)
Total Workload of Machines	Gen/Lin (2012)
Critical Workload of Machines	Yuan/Xu (2013)
Assembly Lines	Relating Articles
Work Overload	Tamura et al. (1999)
Line Stoppages / Duration of Stoppage	Xiaobo/Ohno (1994); Xiaobo/Ohno (1997)
Inventory Costs (Parts and Final Products)	Boysen et al. (2008); Wang/Sarker (2005)
Earliness and Tardiness Costs	Bolat (2003)
Balancing Part Usage	Tamura et al. (1999)
Set Up Costs	Rabbani et al. (2006)
Line Length	Bard et al. (1994)
Makespan / Throughput Time	McMullen (1998)
Worker Displacements	Bard et al. (1994)
Idle Time	Song et al. (2001)
Lead Times	Xiaobo/Ohno (1994)
Costs of Materials	Wang/Sarker (2005)

Source: *own table*

3.2 Grouping

Table 1 already groups the different objectives with regard to the type of production design. Nevertheless, this study is less concerned with production designs than with the effects that the implemen-

tation of specific objectives will have. Thus, groups are constructed according to the effects the contained objectives will have on the company, the involved workers, the products, and the customers. While the construction of the following six groups, plus a seventh group for the residual objectives, might seem arbitrary, their construction resulted from qualitative clustering. The grouping primarily serves to simplify the following discussion. These groups are not absolute, and in some cases, their borders might blur or overlap. The number of groups considered is kept as small as possible while trying to make them internally as homogenous as possible.

A limited number of groups will make it easier for practitioners to select the relevant top level group and, in a second step, consider the most relevant objective for their particular situation.

Group 1 – Makespan related

The first group consists of objectives aimed at minimizing the makespan, lead time, or throughput-times. It therefore also contains the minimization of the total, the mean completion, or the flow time (Benavides & Ritt 2016; McMullen 1998; Pugazhendhi et al. 2004; Sun et al. 2010). Since the length of an assembly line is linked to the lead time (Bard et al. 1994), it is also considered a part of this group. Another objective to be assigned to this group is the minimization of inventory costs for intermediate products (Park & Kim 1999), as the longer an intermediate product is kept in inventory or storage, the longer its lead time will be. Since more changeovers and setups automatically increase a product's lead time, minimizing changeover and setup costs (Grau et al. 1995; Rabbani et al. 2006) are also assigned to this first group. The absence of buffers and the design concept of assembly lines as a continuous flow of products preclude most assembly line objectives from this group. However, in cases where the continuous flow of products is interrupted, for example through the use of line stoppages (Xiaobo & Ohno 1994, 1997), the lead times of the products are impacted. Finally, the total and critical workload of the machines as well as the utilization rate can be assigned to this group if they allow for the accrual of backlogs, which – in consequence – lead to longer lead times.

Group 2 – Workload related

In contrast to the assembly line literature with an inherent focus on Japan (Xiaobo & Ohno 1997), which mainly concentrated on minimizing line stoppages, the US-focused articles were mainly concerned with minimizing work overloads of various types (Bard et al. 1994; Tamura et al. 1999). Worker displacements can be considered a slightly different take on work overloads covered via floater use or line stoppages. The only difference lies in the rigidity of station borders. With rigid station borders, workers cannot leave their station and no displacements occur; the work overloads, however, still need to be covered.

The flip-side to work overloads is idle time (Xiaobo & Ohno 1997), because too much idle time (similar to too much work overload) points to a poorly balanced production run and will put unnecessary stress on the workers. Within job-shop-related literature, the same problem is covered by approaches focusing on the critical workload of the machines (Yuan & Xu 2013). The total workload of the machines can also be assigned to this group if it allows for overutilization and, thus, production backlogs and increasing inventories.

Group 3 – Balanced Production Schedule

Achieving a balanced production schedule (variant-wise or regarding part usage) is one of the major objectives in assembly line scheduling (Boysen et al. 2009), following the assumption that a balanced production schedule will avoid work overloads (Tamura et al. 1999) while the order frequency for parts and intermediate products remains constant. In this regard, it follows the same goal as directly balancing part usage rates. Minimizing parts delivery costs is reflected in the flow store literature by shipping costs when they focus on inputs (Kazemi et al. 2017) and when the assumption is maintained that a machine uses the same parts for each processing of a product, by realizing the average workload of a machine. Focusing on minimizing shipments of parts and intermediate products between plants (Lin et al. 2012), the same idea is present in the resource transition costs or the mileage of vehicles if their focus is on the delivery of parts and not the delivery of final products.

Group 4 – Due date-oriented Goals

This group is similar to the first, but while the first group focused on production time per se, this one focuses on whether the set schedule or an increase in the production makespan results in a delay (Bolat 2003; Chen et al. 2012; Ziaee 2013). For products to become tardy, upper limits for the delivery date need to be set. Thus, minimizing tardiness or lateness costs belongs to this group of objectives. Furthermore, if lower bounds for the delivery date exist, as in an earliest date of delivery, the minimization of the resulting costs can also be assigned to this group. If the focus is on keeping the respective products in stock, the objective belongs to the succeeding group 5. Hence, group 4's objectives relate to group 1's objectives.

Group 5 – Inventory Costs

As discussed in the previous groups that already considered inventory-oriented goals, there are three main types of inventories: inventories of parts (Boysen et al. 2008), inventories of intermediate products, and inventories of final products (Park & Kim 1999; Perret et al. 2022). At this point, intermediate products only refer to those intermediate products produced in the plant itself and stored in buffers

or inventories between stations. The first type of inventory is unavoidable but depends on the use of the parts and the space available on the machines, assembly lines, or in the plant in general. The second type of inventory is null for assembly line designs (Scholl 1999), but can become relevant for flow- and job-shop production. Lastly, inventories of final products only become relevant when push production is implemented or products are processed in orders or batches rather than individually.

Group 6 – Delivery and Shipping Costs

Delivery and shipping costs are the counterpoint to group 3, which focused on parts shipments, whereas this group concentrates on the final products. As such, the minimization of the mileage of vehicles is considered an additional objective that falls into this group (Yan et al. 2022), if the vehicles considered are transporting finished goods.

Group 7 – Other objectives

The objective of maximizing firm income (Doganis et al. 2005) can be seen as an overreaching objective, similar to production efficiency (Fattahi & Fallahi 2010). Thus, both objectives are too broad to be evaluated from a particular sustainability point of view and are excluded from the following analysis.

Having described the six main groups and the seventh residual one, the qualitative clustering as a preliminary step in achieving the objective of this study has been realized. The inherent overarching goal of all objectives, and thus all groups, lies in realizing an increase in output levels or a decrease in costs. Still, they inadvertently target other secondary objectives as well. In the following section, these secondary objectives are discussed for each group with regard to the three pillars of sustainability.

3.3 Sustainability

In the context of the following discussion, the study assumes a definition of sustainability following the three-pillar model as adopted by Basiago (1998) or Purvis et al. (2019), differentiating between ecological, social, and economic sustainability.

Ecological sustainability, herein, is primarily understood as a decrease in resource utilization and harmful emissions. Social sustainability is understood as the living and working conditions of the population in particular and societal well-being in general. Finally, economic sustainability is understood as a positive, stable long-term economic development or, in the production context, as cost reduction in the long term.

In order to find out which sustainability aspects each of the groups addresses in addition to economic sustainability, the groups are discussed individually in turn. As part of this, side effects of implementing certain goals, such as increasing customer satisfaction, are also addressed.

Group 1 – Makespan related

Being able to finish a product earlier translates into an earlier delivery to the customer. The advantage in this regard is reflected in the waiting time for the customers and, thus, in more satisfied customers. Although this does not per se translate into more sustainability, it is not a negligible aspect. Realizing short makespans, however, allows for a shift to pull production, avoiding unnecessary production runs and thus less waste of materials and intermediate products. While pull production may not solve the ‘problem’ of fast fashion, it will at least avoid unnecessary output on the producer’s side (Kiran 2019).

Another advantage related to sustainability in this context is that products are usually paid once they are delivered. While they are kept in stock, they represent bound capital that cannot be used otherwise and is therefore unnecessarily taking up excess inventory space. Combining these aspects, shorter makespans or lead times can at least partially help to avoid or reduce situations like the Chilean Atacama Desert (DW 2021) by reducing overproduction by the manufacturers. In the long run, this would be ecologically beneficial. First and foremost, a focus on smaller inventories will save costs for the companies involved and free up capital that can be invested otherwise. If waste itself is considered an ecological issue, it will again be ecologically sustainable.

It needs to be noted that the proposed objectives, if realized, will lead to developments that carry slight social burdens as well. With less waste occurring and fewer garments dumped into sites like the Atacama Desert, the local population would be deprived of a free source of clothing. Additionally, realizing a leaner production might also reduce the workforce, leading to workers losing their jobs. Even though this is a negative social effect, the higher qualification level of the remaining workers would balance the lost jobs out, and it would therefore still be possible to reach higher ecological sustainability, further evolve the company, and achieve other social goals.

Group 2 – Workload related

A balanced production schedule will have the added benefit of being less stressful not only for the workers involved but also for the implemented machinery. If a machine has to continuously work at maximum capacity, it might result in outages or additional faults in the production runs, which, in turn, would increase the makespan and reduce the quality and, thus, the life-span of products. Aside from consumer preferences, this will play right into the throw-away mentality surrounding fast fashion items. Similarly, stressed workers will be more likely to make mistakes and tend to get sick faster. This effect occurs regardless of whether the overload is achieved through actual or theoretical floater employment or through line stoppages, both of which mean the employee has failed to meet his or her allotted time.

At first glance, idle times, the opposite of work overloads, seem to be beneficial to the workers since they can enjoy some free time. This holds true as long as the idle times are within reasonable bounds and do not alternate with significant work overloads. In the second case, the workers will be stressed by idle times as well, knowing that they will face a work overload. If idle time becomes commonplace, workers may begin to question their necessity or become bored, which will adversely affect their productivity and product quality in the long run. Thus, realizing balanced workloads will directly lead to a socially more sustainable production, indirectly to fewer faults, and consequently to both longer product lifetimes and positive ecological effects.

Group 3 – Balanced Production Schedule

A balanced production sequence helps in realizing two main goals. First, a balanced production sequence will balance workloads at the stations, and all the positive effects mentioned in the previous part will also occur. Second, a balanced production schedule results in even or at least predictable requirements for raw materials and intermediate products (Scholl 1999), realizing a basic requirement for just-in-time production. Sustainability-wise, this will result in better planning of order policies through easier requirement forecasts, translating into better supplier relations. As a result, this can both have a monetary advantage and be considered socially advantageous, especially for the employees involved in contractual work. This means that supply runs of operating materials can be better planned and thus also bundled, which ultimately leads to fewer required runs. In turn, fewer deliveries will mean lower transportation costs but – more importantly – lower fuel and energy requirements and therefore higher ecological sustainability.

Group 4 – Due date-oriented Goals

Due date-oriented goals do not primarily target ecological or social aspects of sustainability, and even economical aspects only matter because of fees related to violations of delivery horizons. Without these fees, the due date-oriented objectives ostensibly result in more satisfied customers.

From a secondary perspective, a production run that avoids tardiness will result in less stressed workers because if workers know they are working on an order that is already late, it will cause significant stress for them (Otten et al. 2019). Consequently, the same effects (even on a smaller scale) might occur as in the case of work overloads. Additionally, considering the delivery of the final product, a production schedule that avoids tardiness, i.e. finishes products as early as possible, will give the company more potential to batch deliveries and save on transportation costs as well as fuel or energy usage. Nevertheless, this advantage comes at the cost of requiring larger inventories for the final products.

Group 5 – Inventory Costs

Realizing a schedule that reduces inventory costs works like the flip side of Group 1's objectives. Minimizing inventory costs will reduce inventory and save inventory space. Due to shorter storage times, it will also lead to reduced lead times.

Depending on the type of product and the duration of storage, the product may deteriorate or, given the short cycle times in the fashion industry, simply go out of style, creating avoidable waste. In particular, if final inventories are kept small, more frequent delivery trips are required, consuming more fuel and energy. Considering that, a reduction in inventories will also free up physical space. Even if these spaces cannot be sold off or repurposed, less energy is required to maintain them, so powering and lighting unnecessary areas would be avoided.

Group 6 – Delivery and Shipping Costs

Minimizing shipment and delivery costs will lead to stronger batching and fewer transports. This will result in lower fuel and energy consumption and a reduction in fixed delivery costs as the company's delivery fleet is downsized. On the downside, realizing fewer shipments and deliveries will force the company to operate larger inventories for parts, intermediate goods, and finished products. Therefore, these objectives operate in opposition to those of Group 5.

Group 7 – Other Objectives

Since the two objectives in this group address overarching aspects, that impact the company as a whole, determining particular sustainability effects does not seem prudent.

Summarizing, it can be seen that a significant inherent potential for more sustainable production lies within most of the objectives currently known and used in mathematical production planning and sequencing. Even though, practitioners might already have realized parts of these inherent potentials, the overview generated in the preceding paragraphs delivers a first comprehensive discussion of the potentials for the three main types of production layouts.

The overview also illustrates that significant goal conflicts exist (groups 1, 5 and 6) which can not be solved in kind. Nevertheless, an optimization approach implementing multiple objectives can find Pareto-optimal solutions from which the decision maker can select the ones that best fit the company's strategy.

Additional practical and theoretical insights to be derived from this study are detailed in the following section.

4 Discussion of Added Values and Implications

Linking scheduling to fashion production and treating it as a possible improvement approach has not been something that has been frequently discussed in recent fashion literature (Perret 2023; Perret et al. 2022). However, it provides great opportunities. Considering the different aspects addressed in the literature and elaborated in the previous chapter, a number of relevant objectives for the improvement of fashion production were presented. The discussion in the second section of this article displayed that the optimization of fashion manufacturing is also possible without the imperative use of high investments in technology or in new machines – thus, a resource-saving improvement approach is definitely feasible. In addition, the advantage of the offered overview lies in its applicability: With the overview presented, manufacturers are equipped with a comprehensive instruction for action that merely needs to be put into a practical setting, i.e. selecting a suitable set of constraints reflecting the concrete production environment. By focusing on reducing sources of waste and working with existing resources, one also has a very company- and operation-oriented improvement methodology at hand. Further, all relevant production steps are included since both employees in manufacturing and logistics as well as possible auxiliary workers (floaters) are considered.

In the past, sustainability in fashion production was widely perceived from the perspective of ecologically and socially questionable conditions and with a lot of criticism (Niinimäki et al. 2020). As a result, previous work has focused less on the potential in terms of sustainable business practices that can originate from the actual production process. Due to savings in working time, liberated employee energy, saved materials (less waste and waste that is harmful to nature), and space that is either returned to nature or at least no longer wasted ‘superfluously’, as well as an optimization of transport routes and logistics (less fuel and energy consumed), the objectives described can enable resource efficiency.

As argued in the previous section, the relationship with customers can probably also be improved by optimizing the production process: In terms of sustainability, this means that increased satisfaction due to shorter waiting times, more flexibility, and most likely better product quality, pay off in the long term because it increases loyalty and thus, secures future sales and jobs. All of these aspects, consequently, result in both ecological and social sustainability as can be seen from discussions concerning ESG criteria (Câmara & Morais 2022). Further, since financial savings and reductions in fixed costs are highly conceivable, the application of the objectives discussed can also lead to economic sustainability, which – in turn – can be used for advanced sustainability projects, aligned with the corporate strategy and orientation. Examples of these include, but are not limited to, training and education opportunities for employees, charity projects, infrastructural projects for the local population, or environmental protection projects with regard to the areas negatively affected by the clothing industry.

Based on the above, there are also some aspects that are required from the management side when it comes to successfully implementing optimization projects in fashion productions and achieving the described objectives: Firstly, a corporate culture that is open-minded when it comes to change processes is strongly required. Any change process in companies rarely goes along without resistance from the employees' side, so transparent communication and the involvement of the workers are absolutely imperative if the changes are to be lastingly successful (Hines & Bruce 2007). In addition, regular analyses and an efficient information exchange system are essential, as this is the only way to regularly identify waste sources, improvement potentials, and actual time and resource allocations. Furthermore, optimization projects can only be managed efficiently if responsibilities and action plans are clearly defined, including targets and key performance indicators (Towers & McLoughlin 2005). Regular checks are also mandatory, as otherwise there is no long-term improvement that is both measurable and visible. But dealing with optimization also requires a responsible course of action: Reducing the number of employees, which might come along with some optimization objectives, could pose a new problem itself as many of the workers in fashion are highly dependent on their jobs, especially in the low-wage countries where fashion is often produced. Therefore, in addition to assistance in finding other positions, the manufacturer could offer training or educational programs as well as coaching for different manufacturing steps in order to keep the employees in the company. Further, at some production sites, some employees have generated their income from either free clothing or from the further processing of the respective textile waste. While optimizing production and rationalizing these income sources, production companies would need to take responsibility and offer alternatives.

Considering fashion manufacturing beyond the mere production process, it becomes evident that flexibility and an efficient supply chain partnership network would be necessary in order to optimize production above the limits of the company's own capabilities. This comes with the managerial challenge of long-term thinking, the commitment to suppliers, and their support in their own optimization. But, at the same time, it is also about being prepared and entering into new partnerships, expanding the existing network (Rafi-Ul-Shan et al. 2018). Realizing a more optimal production process frees resources and saves money that can finance these endeavors. Although this was not the explicit focus of this paper, it is nevertheless worth mentioning in this context.

5 Outlook and Limitations

In summary, it can be concluded that the garment industry is an extremely relevant industry that has a great influence on the environment – from an economic, but above all, from an ecological and social perspective (Niinimäki et al. 2020). Since garments are a basic necessity, the industry also holds a great deal of responsibility which (in many cases) is not born as such. Although the COVID-19 pandemic has triggered some changes and raised important issues for the industry, such as seasonless fashion or the inclusion of more sustainable pieces, as of 2022 and 2023, it has already become apparent that some

brands are falling back into former habits. For this reason, it is more relevant than ever to make sustainability lucrative and therefore interesting to apply to an industry as financially motivated as the fashion industry. Ultimately, optimizing production towards a more sustainable way of working is also about questioning established business models. For the fashion sector, this means a switch from fast to slow fashion, i.e. slow the cycles down and produce fewer collections in order to both reduce the pressure on production workers and cause less damage to the environment. But at the same time, it also means improving the already existing business models. Therefore, the meta-study in this paper delivered different objectives and accompanied optimization ideas that are applicable to various production formats for garment manufacturing while assuming the resources already available. It also introduced a different approach to thinking about production optimization because it displayed to what extent production optimization might also lead to more sustainable business activities (social and ecological) and how it can release potentials for sustainability activities and projects, beyond the bounds of material use, recycling, and cycle slowdown.

But it is not without limitations: The aspects described only cover one part of the value chain, namely the manufacturing process. However, depending on the production, distribution, and sales locations, a value chain in the garment industry can consist of various internationally connected and multi-layered stages, making it very complex (Backs et al. 2021). The process mostly starts with the cultivation of raw materials such as cotton or the manufacture of artificial or synthetic fibers and proceeds with the subsequent production of yarn and textile fabrics. These steps are usually processed in Southeast Asia. After the garments have been produced, they are transported to distribution centers or warehouses, followed by the retail steps (Backs et al. 2021; Niinimäki et al. 2020). The actual garment production is therefore only one small part of a whole chain, which is why further research should address other parts of the value chain in order to consider a more holistic optimization approach. In addition, aspects such as minimum and maximum order quantities, delivery times, and specific regulations (e.g. import and export constraints) should be included in further research. Moreover, the article does not offer a specialization on a specific case and does not apply the proposed optimization ideas to a 'real' company, including an evaluation of the employees. On the one hand, this gave the possibility of obtaining generalizable findings but, on the other hand, severely limited the accuracy of the results. Hence, it is suggested that researchers and practitioners apply the proposed ideas, i.e. put a distinct price tag on the proposed ideas, and also investigate the company's staff perspective as they play a crucial role in implementing new approaches.

Furthermore, only limited attention was paid to the changes that the pandemic has already brought or will bring to the production process (but also to the entire supply chain) of fashion companies in the future. In this context, researchers discussed the topic of 'near and re-shoring', which is the relocation of production to a site that is close to the respective company location. The ideas described strive to provide easy-to-implement options that do not necessarily require high investments. However, this

only provides companies with limited optimization options that can be beneficially extended with other approaches: For example, Bertola and Teunissen (2018) present ideas on how the use of technology (data analysis, 3D printers, and smart factory appliances) can help to optimize fashion production. Researchers are therefore reminded to consider various dimensions and approaches from different industries when it comes to achieving optimization in garment manufacturing. Moreover, while COVID-19 has had a negative impact on fashion companies, some practitioners believe at the same time that it might have a particularly positive impact on the garment industry as it strongly encouraged the use of robots and the associated automation in China, for example. Perhaps this will soon become the case in the global garment industry because, where people still work manually, the demand to always become faster and better can no longer be sustained at some point.

Finally, the article intentionally does not focus on any particular optimization method, use case or case study. As in the past decades, optimization techniques will improve in the coming years, not only methodologically but also with regard to computational power; e.g., due to the availability of quantum computing or the increased use of artificial intelligence to solve even complex optimization problems. While the techniques and methods will change, and even production layouts may vary over time, asking the right and most relevant questions will remain at the core of all practical applications of optimization.

While a practical use case might illustrate the theoretical aspects discussed herein, it needs to be considered that actual data, but even more so the implemented optimization approaches, are usually treated as strictly proprietary. With regard to the optimization of the production of automobiles, the sector most prominently featured in the literature, the competition by PSA launched in the 2000s, where they made their algorithm public, has been the only instance where a company shared this type of knowledge.

Another solution to this conundrum might be an in-depth qualitative interview-based study, using the opinions of experts to evaluate the ideas put forth in this study.

Summarizing, this article provides not just a solution to a single current problem but an overview of the major questions that might rightfully be asked and evaluates their relevance with regard to sustainability, making it more timeless than focusing on a single use case could ever be.

References

- Ait-Alla, A.; Teucke, M.; Lütjen, M.; Beheshti-Kashi, S.; Karimi, H. R. (2014): Robust production planning in fashion apparel industry under demand uncertainty via conditional value at risk. In: *Mathematical Problems in Engineering* (2014).
- Al-Anzi, F. S.; Allahverdi, A. (2007): A Self-Adaptive Differential Evolution Heuristic for Two-Stage Assembly Scheduling Problem to Minimize Maximum Lateness with Setup Times. In: *European Journal of Operational Research*, 182. (2007), No. 1, pp. 80–94.

- Al-Anzi, F. S.; Allahverdi, A. (2013): An Artificial Immune System Heuristic for Two-Stage Multi-Machine Assembly Scheduling Problem to Minimize Total Completion Time. In: *Journal of Manufacturing Systems*, 32. (2013), No. 4, pp. 825–830.
- Arvidsson, A.; Malossi, G.; Naro, S. (2010): Passionate work? Labor conditions in the Milan fashion industry. In: *Journal for Cultural Research*, 14. (2010), No. 3, pp. 295–309.
- Backs, S.; Jahnke, H.; Lüpke, L.; Stücken, M.; Stummer, C. (2021): Traditional versus fast fashion supply chains in the apparel industry: an agent-based simulation approach. In: *Annals of Operations Research*, 305. (2021), pp. 487–512.
- Bard, J. F.; Shtub, A.; Joshi, S. B. (1994): Sequencing mixed-model assembly lines to level parts usage and minimize line length. In: *The International Journal of Production Research*, 32. (1994), No. 10, pp. 2431–2454.
- Barnes, L.; Lea-Greenwood, G. (2006): Fast fashioning the supply chain: Shaping the research agenda. In: *Journal of Fashion Marketing and Management: An International Journal*, 10. (2006), No. 3, pp. 259–271.
- Basiago, A. D. (1998): Economic, social, and environmental sustainability in development theory and urban planning practice. In: *Environmentalist*, 19. (1998), pp. 145–161.
- Battini, D.; Faccio, M.; Persona, A.; Sgarbossa, F. (2009): Balancing–sequencing procedure for a mixed model assembly system in case of finite buffer capacity. In: *The International Journal of Advanced Manufacturing Technology*, 44. (2009), pp. 345–359.
- BCG (2019): Pulse of the Fashion Industry (https://web-assets.bcg.com/img-src/Pulse-of-the-Fashion-Industry2019_tcm9-237791.pdf). Accessed on 01.02.2022.
- Benavides, A. J.; Ritt, M. (2016): Two simple and effective heuristics for minimizing the makespan in non-permutation flow shops. In: *Computers & Operations Research*, 66. (2016), pp. 160–169.
- Bertola, P.; Teunissen, J. (2018): Fashion 4.0. Innovating fashion industry through digital transformation. In: *Research Journal of Textile and Apparel*, 22. (2018), No. 4, pp. 352–369.
- Bevilacqua, M.; Ciarapica, F. E.; Crosta, A.; Mazzuto, G.; Paciarotti, C. (2013): Designing an efficient production system: A case study of a clothing company. In: *International Journal of Engineering Business Management*, 5. (2013).
- Blazewicz, J.; Ecker, K. H.; Pesch, E.; Schmidt, G.; Sterna, M.; Weglarz, J. (2019): *Handbook on Scheduling. From Theory to Practice*. 2.th ed., Wiesbaden: Springer.
- Blocher, J. D.; Chhajer, D. (2008): Minimizing Customer Order Lead-Time in a Two-Stage Assembly Supply Chain. In: *Annals of Operations Research*, 161. (2008), No. 1, pp. 25–52.
- BoF & McKinsey (2021): The State of Fashion 2020 (<https://www.mckinsey.com/~media/mckinsey/industries/retail/our%20insights/the%20state%20of%20fashion%202020%20navigating%20uncertainty/the-state-of-fashion-2020-final.pdf>). Accessed on 01.02.2022.
- Bolat, A. (2003): A mathematical model for selecting mixed models with due dates. In: *International Journal of Production Research*, 41. (2003), No. 5, pp. 897–918.
- Boysen, N.; Fliedner, M.; Scholl, A. (2008): Sequencing mixed-model assembly lines to minimize part inventory cost. In: *OR Spectrum*, 30. (2008), No. 3, pp. 611–633.

- Boysen, N.; Fliedner, M.; Scholl, A. (2009): Sequencing mixed-model assembly lines: Survey, classification and model critique. In: *European Journal of Operational Research*, 192. (2009), No. 2, pp. 349–373.
- Bruce, M.; Daly, L.; Towers, N. (2004): Lean or agile: A solution for supply chain management in the textiles and clothing industry? In: *International Journal of Operations & Production Management*, 24. (2004), No. 2, pp. 151–170.
- Brydges, T.; Retamal, M.; Hanlon, M. (2020): Will COVID-19 support the transition to a more sustainable fashion industry? In: *Sustainability: Science, Practice and Policy*, 16. (2020), No. 1, pp. 298–308.
- Cachon, G. P.; Swinney, R. (2011): The value of fast fashion: Quick response, enhanced design, and strategic consumer behavior. In: *Management Science*, 57. (2011), No. 4, pp. 778–795.
- Câmara, P.; Morais, F. (eds.) (2022): *The Palgrave Handbook of ESG and Corporate Governance*. London: Palgrave Macmillan.
- Camargo, L. R.; Farias Pereira, S. C.; Santiago, Scarpin, M.R. (2020): Fast and ultra-fast fashion supply chain management: An exploratory research. In: *International Journal of Retail & Distribution Management*, 48. (2020), No. 6, pp. 537–553.
- Chaudhry, I. A.; Khan, A. A. (2016): A research survey: review of flexible job shop scheduling techniques. In: *International Transactions in Operational Research*, 23. (2016), No. 3, pp. 551–591.
- Chen, J.; Chen, F. (2008): Adaptive scheduling and tool flow control in flexible job shops. In: *International Journal of Production Research*, 46. (2008), No. 15, pp. 4035–4059.
- Chen, J. C.; Wu, C.-C.; Chen, C.-W.; Chen, K.-H. (2012): Flexible job shop scheduling with parallel machines using genetic algorithm and grouping genetic algorithm. In: *Expert Systems with Applications*, 39. (2012), No. 11, pp. 10016–10021.
- Choi, T.-M.; Hui, C.-L.; Yu, Y. (2014): *Intelligent Fashion Forecasting Systems: Models and Applications*. Wiesbaden: Springer.
- Christopher, M.; Lowson, R.; Peck, H. (2004): Creating agile supply chains in the fashion industry. In: *International Journal of Retail & Distribution Management*, 32. (2004), No. 8, pp. 367–376.
- de Oliveira Neto, G. C.; Ferreira Correia, J. M.; Silva, P. C.; de Oliveira Sanches, A. G.; Lucato, W. C. (2019): Cleaner Production in the Textile Industry and its Relationship to Sustainable Development Goals. In: *Journal of Cleaner Production*, 228. (2019), pp. 1514–1525.
- Debnath, S. (2016): Pineapple Lead Fibre – A Sustainable Luxury and Industrial Textiles. In: Gardetti, M. A.; Muthu, S. S. (eds.): *Handbook of Sustainable Luxury Textiles and Fashion*. Volume 2. Heidelberg: Springer, pp. 35–50.
- Defersha, F. M.; Mohebalizadehgashti (2018): Simultaneous balancing, sequencing, and workstation planning for a mixed model manual assembly line using hybrid genetic algorithm. In: *Computers & Industrial Engineering*, 119. (2018), pp. 370–387.
- Dhamala, T. N.; Kubiak, W. (2005): A brief survey of just-in-time sequencing for mixed-model systems. In: *International Journal of Operational Research*, 2. (2005), No. 2, pp. 38–47.

- Doganis, P.; Sarimveis, H.; Bafas, G.; Koufos, D. (2005): An MILP model for optimal scheduling of the lubricant production plant. In: *Chemical Engineering Communications*, 192. (2005), No. 8, pp. 1067–1084.
- DW (2021): Fast fashion on one-way trip to Atacama Desert dump (<https://www.dw.com/en/fast-fashion-on-one-way-trip-to-atacama-desert-dump/av-60133560>). Accessed on 01.02.2022.
- Eira, R.; Maia, L. C.; Alves, A. C.; Leao, C. P. (2015): An Initiation of a lean Journey in a Clothing Company. In: Gomes, J.; Meguid, S. A. (eds.): *M2D2015: Proceedings of the 6th International Conference on Mechanics and Materials in Design*. Porto: INEGI-FEUP, pp. 1349–1358.
- Fani, V. (2018): *A Simulation Optimization Framework for Production Planning and Control in the Fashion Industry*; Università degli studi Firenze (ed.), Florence.
- Fattahi, P.; Fallahi, A. (2010): Dynamic scheduling in flexible job shop systems by considering simultaneously efficiency and stability. In: *CIRP Journal of Manufacturing Science and Technology*, 2. (2010), No. 2, pp. 114–123.
- Gen, M.; Lin, L. (2012): Multiobjective genetic algorithm for scheduling problems in manufacturing systems. In: *Industrial Engineering & Management Systems*, 11. (2012), No. 4, pp. 310–330.
- Grau, R.; Espuna, A.; Puigjaner, L. (1995): Environmental considerations in batch production scheduling. In: *Computers & Chemical Engineering*, 19. (1995), pp. 651–656.
- Gwilt, A. (2020): *A Practical Guide to Sustainable Fashion*. 2nd ed., London: Bloomsbury Publishing.
- Hines, T.; Bruce, M. (2007): *Fashion Marketing. Contemporary Issues*. 2nd ed., Oxford: Butterworth-Heinemann.
- Kazemi, H.; Mazdeh, M. M.; Rostami, M. (2017): The Two Stage Assembly Flow-Shop Scheduling Problem with Batching and Delivery. In: *Engineering Applications of Artificial Intelligence*, 63. (2017), pp. 98–107.
- Kim, M. (2021): Green is the New Black: The Effects of COVID-19 on the Fashion Industry's Need for Sustainability (Joseph Wharton Scholars) (https://repository.upenn.edu/joseph_wharton_scholars/108). Accessed on 01.02.2022.
- Kiran, D. R. (2019): *Production Planning and Control*. Oxford: Butterworth-Heinemann.
- Komaki, G. M.; Sheikh, S.; Malakooti, B. (2019): Flow shop scheduling problems with assembly operations: A review and new trends. In: *International Journal of Production Research*, 57. (2019), No. 10, pp. 2926–2955.
- Kulsum, U. (2020): Sustainable Fashion as The Early Awakening of the Clothing Industry Post Corona Pandemic. In: *International Journal of Social Science and Business*, 4. (2020), No. 3, pp. 422–429.
- Lan, M.; Xu, T.-R.; Peng, L. (2010): Solving flexible multi-objective JSP problem using a improved genetic algorithm. In: *Journal of Software*, 5. (2010), No. 10, pp. 1107–1113.
- Lin, L.; Hao, X. C.; Gen, M.; Jo, J. B. (2012): Network modeling and evolutionary optimization for scheduling in manufacturing. In: *Journal of Intelligent Manufacturing*, 26. (2012), No. 6, pp. 2237–2253.
- Liu, N.; Chow, P.-S.; Zhao, H. (2020): Challenges and Critical Successful Factors for Apparel Mass Customization Operations: Recent Development and Case Study. In: *Annals of Operations Research*, 291. (2020), No. 1, pp. 531–563.

- Lopes, T. C.; Sikora, C.; Michels, A. S.; Magatao, L. (2020): An iterative decomposition for asynchronous mixed-model assembly lines: combining balancing, sequencing, and buffer allocation. In: *International Journal of Production Research*, 58. (2020), No. 2, pp. 615–630.
- McMullen, P. R. (1998): JIT sequencing for mixed-model assembly lines with setups using tabu search. In: *Production Planning & Control*, 5. (1998), No. 5, pp. 504–510.
- Nie, L.; Gao, L.; Li, P.; Li, X. (2013): A GEP-based reactive scheduling policies constructing approach for dynamic flexible job shop scheduling problem with job release dates. In: *Journal of Intelligent Manufacturing*, 24. (2013), No. 4, pp. 763–774.
- Niinimäki, K.; Peters, G.; Dahlbo, H.; Perry, P.; Rissanen, T.; Gwilt, A. (2020): The environmental price of fast fashion. In: *Nature Reviews Earth & Environment*, 1. (2020), No. 4, pp. 189–200.
- Otten, M.; Braaksma, A.; Boucherie, R. J. (2019): Minimizing Earliness/Tardiness costs on multiple machines with an application to surgery scheduling. In: *Operations Research for Health Care*, 22. (2019), p. 100194.
- Pal, R.; Gander, J. (2018): Modelling environmental value: An examination of sustainable business models within the fashion industry. In: *Journal of Cleaner Production*, 184. (2018), pp. 251–263.
- Park, M. W.; Kim, Y. D. (1999): A Heuristic Algorithm for a Production Planning Problem in an Assembly System. In: *Journal of Operational Research Society*, 50. (1999), No. 2, pp. 132–143.
- Parveen, S.; Ullah, H. (2010): Review on job-shop and flow-shop scheduling using multi criteria decision making. In: *Journal of Mechanical Engineering*, 41. (2010), No. 2, pp. 130–146.
- Perret, J. K. (2022): A simultaneous balancing and sequencing algorithm to plan assembly lines in the fashion industry. In: *Research Journal of Textile and Apparel*, vol. Online First (2022).
- Perret, J. K.; Schuck, K. A.; Hitzegrad, C. (2022): Production Scheduling of Personalized Fashion Goods in a Mass Customization Environment. In: *Sustainability*, 14. (2022), No. 1, p. 538.
- Perry, P. (2017): Garments without guilt? In: Gardetti, M. A.; Torres, A. L. (eds.): *Sustainability in Fashion and Textiles*. London: Routledge, pp. 294–307.
- Perry, P.; Towers, N. (2013): Conceptual framework development: CSR implementation in fashion supply chains. In: *International Journal of Physical Distribution & Logistics Management*, 43. (2013), No. 5/6, pp. 478–500.
- Pugazhendhi, S.; Thiagarajan, S.; Rajendran, C. (2004): Relative performance evaluation of permutation and non-permutation schedules in flowline-based manufacturing systems with flowtime objective. In: *The International Journal of Advanced Manufacturing Technology*, 23. (2004), No. 11–12, pp. 820–830.
- Purvis, B.; Mao, Y.; Robinson, D. (2019): Three pillars of sustainability: In search of conceptual origins. In: *Sustainability Science*, 14. (2019), pp. 681–695.
- Rabbani, M.; Rahimi-Vahed, A.; Javadi, B.; Tavakkoli-Moghaddam, R. (2006): A New Approach for Mixed-Model Assembly Line Sequencing. In: Waldmann, K.-H.; Stocker, U. M. (eds.): *Operations Research Proceedings 2006*. Wiesbaden: Springer, pp. 169–174.
- Rafi-Ul-Shan, P. M.; Grant, D. B.; Perry, P. (2017): Coopetition in Fashion Supply Chains. In: *Global Fashion Management Conference*, (2017), pp. 176–180.

- Rafi-Ul-Shan, P. M.; Grant, D. B.; Perry, P.; Ahmed, S. (2018): Relationship between sustainability and risk management in fashion supply chains: A systematic literature review. In: *International Journal of Retail & Distribution Management*, 46. (2018), No. 5, pp. 466–486.
- Rossit, D. A.; Tohmé, F.; Frutos, M. (2018): The non-permutation flow-shop scheduling problem: A literature review. In: *Omega*, 77. (2018), pp. 143–153.
- Scholl, A. (1999): *Balancing and Sequencing of Assembly Lines*. Heidelberg: Springer.
- Song, D. P.; Earl, C. F.; Hicks, C. (2001): Stage due date planning for multistage assembly systems. In: *International Journal of Production Research*, 39. (2001), No. 9, pp. 1943–1954.
- Sumarlia, E.; Usmanova, K.; Fauziyah, F.; Mousa, K. (2021): Managing the risks in the clothing supply chain considering the coronavirus pandemic. In: *Operations and Supply Chain Management: An International Journal*, 14. (2021), No. 4, pp. 576–587.
- Sun, W.; Pan, Y.; Lu, X.; Ma, Q. (2010): Research on flexible job-shop scheduling problem based on a modified genetic algorithm. In: *Journal of Mechanical Science and Technology*, 24. (2010), No. 10, pp. 2119–2125.
- Tamura, T.; Long, H.; Ohno, K. (1999): A sequencing problem to level part usage rates and work loads for a mixed-model assembly line with a bypass subline. In: *International Journal of Production Economics*, 60. (1999), pp. 557–564.
- Towers, N.; McLoughlin, J. (2005): Effective Total Quality Management in the textile fashion retail supply chain: A pilot survey of the UK textile manufacturers. In: *The Journal of the Textile Institute*, 96. (2005), No. 2, pp. 87–92.
- UNEP (2022): Putting the brakes on fast fashion (<https://www.unep.org/news-and-stories/story/putting-brakes-fast-fashion>). Accessed on 01.02.2022.
- Vilcot, G.; Billaut, J.-C. (2011): A tabu search algorithm for solving a multicriteria flexible job shop scheduling problem. In: *International Journal of Production Research*, 49. (2011), No. 23, pp. 6963–6980.
- Wang, S.; Sarker, B. R. (2005): An assembly-type supply chain system controlled by kanbans under a just-in-time delivery policy. In: *European Journal of Operational Research*, 162. (2005), No. 1, pp. 153–172.
- Xiaobo, Z.; Ohno, K. (1994): A sequencing problem for a mixed-model assembly line in a JIT production system. In: *Computers & Industrial Engineering*, 27. (1994), No. 1-4, pp. 71–74.
- Xiaobo, Z.; Ohno, K. (1997): Algorithms for sequencing mixed models on an assembly line in a JIT production system. In: *Computers & Industrial Engineering*, 32. (1997), No. 1, pp. 47–56.
- Xu, Y.; Thomassey, S.; Zeng, X. (2020): Garment mass customization methods for the cutting-related processes. In: *Textile Research Journal*, 9. (2020), No. 7-8, pp. 802–819.
- Yan, H.; Du, X.; Xu, L.; Xu, S.; Zhang, Y. et al. (2022): Toward intelligent clothes manufacturing: a systematic method for static and dynamic task allocation by genetic optimization. In: *Neural Computing and Applications*, 34. (2022), No. 10, pp. 7881–7897.
- Yang, J.-H.; Kincade, D. H.; Chen-Yu, J.H. (2015): Type of Apparel Mass Customization and Levels of Modularity and Variety: Application of the Theory of Inventive Problem Solving. In: *Clothing and Textiles Research Journal*, 33. (2015), No. 3, pp. 199–212.

- Yeung, H.-T.; Choi, T.-M.; Chiu, C.-H. (2010): Innovative Mass Customization in the Fashion Industry. In: Cheng, T.; Choi, T.-M. (eds.): Innovative Quick Response Programs in Logistics and Supply Chain Management. Heidelberg: Springer (International Handbooks on Information Systems), pp. 423–454.
- Yuan, Y.; Xu, H. (2013): Multiobjective flexible job shop scheduling using memetic algorithms. In: IEEE Transactions on Automation Science and Engineering, 12. (2013), No. 1, pp. 1–18.
- Zhang, J.; Ding, G.; Zou, Y.; Qin, S.; Fu, J. (2019): Review of job shop scheduling research and its new perspectives under Industry 4.0. In: Journal of Intelligent Manufacturing, 30. (2019), No. 4, pp. 1809–1830.
- Ziaee, M. (2013): General flowshop scheduling problem with the sequence dependent setup times: A heuristic approach. In: Information Sciences, 25. (2013), No. 126–135.

Authors

Katharina A. **Schuck**, M.A. finished her post-graduate studies in Luxury, Fashion and Sales Management at the International School of Management, German University of Applied Sciences. She holds a Bachelor in Business Administration from the Hochschule Fresenius in Düsseldorf. Her research interests include the luxury and the fashion industry with a focus on general marketing in particular brand management and consumer behaviour as well as topics of manufacturing and change management.



Prof. Dr. Jens K. **Perret** studied business mathematics and economics and received his PhD in economics from the Bergische Universität Wuppertal. Mr. Perret worked at the European Institute for International Economic Relations in Wuppertal. Since 2016, he holds a professorship in Economics and Statistics at the International School of Management, German University of Applied Sciences. His research interests include the use of data analytics in marketing and aspects of the knowledge society.



Gericke, Jens; Mehn, Audrey; Rommel, Kai

Kundenpräferenzen für Bio-Lebensmittel in deutschen Lebensmittelläden. Ergebnisse eines Discrete-Choice Experiments

Abstract

Hintergrund: Diese Studie leistet einen Beitrag zur Erforschung des Verbraucherverhaltens im Lebensmittelsektor und untersucht die Determinanten der Zahlungsbereitschaft für Bio-Produkte wie Milch, Fleisch, Äpfel, Kartoffeln und Wein, die in deutschen Märkten für den schnell-drehenden Konsum dauerhaft verfügbar sind. Ziel der Studie ist es, die Zahlungsbereitschaft zu ermitteln und den Einfluss ausgewählter sozioökonomischer Variablen auf die Präferenzstruktur aufzuzeigen.

Design/Methodik/Vorgehensweise: Die Autoren wenden ein Discrete Choice Experiment an, um die Zahlungsbereitschaft und sozioökonomische Determinanten valide zu bestimmen. Es wurden 400 Interviews mit jeweils acht Wahlkarten durchgeführt, was zu 3.200 Datensätzen führte. Die Verwendung eines Multinomial- und eines Latent-Class-Modells liefert Aussagen über die Heterogenität der Variablen innerhalb und zwischen den Klassen, indem Klassen auf der Grundlage statistisch gewichteter soziodemografischer Merkmale gebildet werden.

Ergebnisse: Die Ergebnisse zeigen signifikante Einflüsse der meisten Merkmalsausprägungen auf die Nutzenfunktion in beiden Modellen und eine stark heterogene Verteilung der Zahlungsbereitschaft zwischen den beiden gebildeten Klassen.

Originalität/Wert: Die Ergebnisse beleuchten nach Kenntnis der Autoren den deutschen Markt für intensiv genutzte Bioprodukte und können zur Optimierung von Marketingkonzepten für den Lebensmitteleinzelhandel speziell auf dem deutschen Markt genutzt werden.

Keywords:

Discrete Choice Experiment, Bio-Lebensmittel, Zahlungsbereitschaft, strategisches Marketing

JEL: D12, D47, M31

1 Einleitung

1.1 Der deutsche Markt für Bio-Lebensmittel

Der Markt für Bio-Lebensmittel in Deutschland ist von 4,1 % im Jahr 2011 auf 6,8 % im Jahr 2021 gewachsen. Allerdings ist die Wachstumsrate im Jahr 2021 eine der niedrigsten seit 2010 (Statista 2022a, b) und im Jahr 2022 ist der Umsatz um etwa 3,5 % gesunken (Statista 2023). Verbraucher kaufen Bio-Produkte vor allem in Supermärkten (91 %), Discounter (71 %), Bäckereien (64 %) und in verschiedenen Lebensmittelgeschäften. Rund 58 % kaufen Bio-Produkte in Biomärkten oder direkt beim Erzeuger (BMEL 2022). Durch externe Faktoren wie die Inflation, die COVID19-Pandemie und Engpässe in der Lieferkette steht der Markt für Bio-Lebensmittel unter besonderem Druck, da die Bereitschaft der Kunden, für Bio-Qualität einen höheren Preis zu zahlen, mit steigenden Ausgaben für dieses Segment im Vergleich zu konventionellen Produkten abnimmt. Seit Beginn der Pandemie ist das Interesse der Verbraucher an Bio-Produkten gestiegen. Im Jahr 2022 gaben rund 82 % der Befragten an, Bio-Lebensmittel häufig oder gelegentlich zu kaufen, im Jahr 2020 waren es noch 72 % (BMEL 2023). Das Wachstumspotenzial in diesem Marktsegment wird von der Nachfrageseite bestimmt, da die Grenzkosten der Anbieter nur sehr begrenzt gesenkt werden können. Dies liegt vor allem an den natürlichen Restriktionen der industriellen landwirtschaftlichen Produktion und an der Optimierung der Grenzkostenkurven, die in den letzten Jahren aufgrund des zunehmenden Wettbewerbs auf weitgehend gesättigten Märkten bereits stattgefunden hat (BOELW 2022).

1.2 Präferenzstrukturen für Bio-Lebensmittel

Die Nachfrage auf dem Markt für Bio-Lebensmittel weist eine ausgeprägte Preissensibilität auf (Meyering 2018). Die Auswirkungen von Kostensteigerungen auf das Verbraucherverhalten und die Auswirkungen von sich verändernden Preisrelationen sind wesentliche Bestimmungsfaktoren für die zukünftigen Wachstumspfade von Bio. Ein deutlicher Nachfragerückgang in Verbindung mit einem Umsatzrückgang im Lebensmitteleinzelhandel wird zu einer Straffung des Sortiments führen. Dies wirkt sich auf die gesamte Lieferkette aus, da die Lebensmittelläden in Deutschland aufgrund ihrer starken Marktposition die notwendigen Preissenkungen weitergeben können, um die Nachfrage bei den Erzeugern anzukurbeln. Um diese Effekte zu messen und die Entwicklungspfade in diesem Marktsegment stabil zu prognostizieren, ist es notwendig, die Präferenzen für einzelne Produktmerkmale zu ermitteln (Bazzani et al. 2017; Marreiros 2021). Diese Daten können genutzt werden, um spezifische Kundenaktionen zu isolieren und zu analysieren.

Das Ziel dieser Studie ist es, Präferenzen für Bio-Lebensmittel in deutschen Lebensmittelgeschäften zu identifizieren. Wir verwenden ein Discrete Choice Experiment (DCE) zur Schätzung der marginalen Zahlungsbereitschaft (ZB) für einzelne Attribute und soziodemografische Determinanten dieser Werte. Nach unserer Kenntnis haben in den letzten Jahren nur sehr wenige Studien DCE verwendet, um dieses deutsche Marktsegment zu untersuchen. Der Rest dieses Beitrags ist wie folgt gegliedert. Abschnitt 2 enthält Hintergrundinformationen auf der Grundlage der relevanten Studien. Abschnitt 3 beschreibt die Methodik, und Abschnitt 4 stellt das Modell und die Ergebnisse vor. In Abschnitt 5 werden die Ergebnisse diskutiert und Empfehlungen und Maßnahmen für den Lebensmitteleinzelhandel abgeleitet.

2 Marktsituation und Stand der Forschung

2.1 Preisdruck auf dem deutschen Markt für Bio-Lebensmittel

Die steigende Inflationsrate in Deutschland, insbesondere bei Lebensmitteln und Energie, und die wahrgenommene Bedrohung der Versorgungssicherheit können dazu führen, dass der bisherige Trend stetig steigender Wachstumsraten für Bio-Lebensmittel zu Ende geht und zu einem Verlust von Marktanteilen führt. Ziel der Untersuchung ist es, herauszufinden, inwieweit die Verbraucher auf diese veränderten Rahmenbedingungen reagieren und ihr bisheriges Kaufverhalten im Bereich Bio-Lebensmittel anpassen. Mehrere Studien zu den Präferenzen der Verbraucher für Bio-Lebensmittel zeigen eine Präferenzheterogenität, die durch unterschiedliche Einflüsse von Produktattributen und soziodemografischen Variablen verursacht wird (z. B. Paul/Rana 2012; Asante-Addo/Weible 2019). Die Einflussfaktoren sind der erwartete Gesundheitsnutzen, die Bildung, die Nutzeneffekte von Bio-Lebensmitteln und die Produktverfügbarkeit. Diese Faktoren haben einen positiven Einfluss auf die ZB für Bio-Produkte (Paul/Rana 2012).

2.2 Wie verschiedene Produkteigenschaften das Kaufverhalten beeinflussen

Auch Persönlichkeitsmerkmale bestimmen das Kaufverhalten. Bazzani et al. (2017) fanden einen positiven Effekt von Verträglichkeit und einen negativen Effekt von Extraversion auf die individuelle ZB für lokal erzeugtes und biologisches Apfelmus. Studien zur Marktsegmentierung zeigen unterschiedliche Einstellungen zu den Attributen der Regionalität von Bio-Produkten. Demografische und verhaltensbezogene Variablen sowie Wissen haben stärkere Auswirkungen auf das Konsumverhalten als die Einstellung zu Bio-Produkten (Marreiros 2021). Dies unterstreicht, dass die Marktsegmentierung neben dem Kaufverhalten und der Kaufhäufigkeit Kenntnisse über soziodemographische Determinanten der Präferenzen für einzelne Attribute von Bio-Produkten erfordert. Lokale und im Inland erzeugte Lebensmittel und Produkte mit Qualitätssiegeln haben positive Auswirkungen auf die individuelle Nutzen-

funktion (Meyerding 2018). Auch die Rückverfolgbarkeit bestimmt das Konsumverhalten auf den Lebensmittelmärkten, wenn die Verbraucher auf Attribute wie Informationen über den Einsatz von Pestiziden und Düngemitteln Wert legen (Kabir et al. 2023). Diese Präferenz wird in unserer Studie durch die ZB für Bio-Siegel dargestellt. Die Ergebnisse zu Gütesiegeln und lokaler Produktion zeigen deutliche Unterschiede im Einfluss der einzelnen Merkmale auf die ZB. Dies wirft die Frage auf, ob lokale Erhebungen für einzelne Produkte, wie die von Bazzani et al. (2017) in Italien, geeignet sind, um ein System zur Beschreibung des Einflusses von soziodemografischen Merkmalen und Einstellungen auf die Ausrichtung und Positionierung von Bio-Segmenten zu entwickeln. Die Verwendung von Gütesiegeln spielt ebenfalls eine wichtige Rolle im Kaufprozess, allerdings bei einem geringen Bekanntheitsgrad von Bio-Siegeln (Meyerding 2018). Der Zusammenhang zwischen dem Konsum von Bio-Lebensmitteln und dem individuellen Gesundheitszustand ist ein weiterer Forschungsgegenstand in der Literatur. Entscheidend ist dabei die individuell wahrgenommene Definition von Bio-Lebensmitteln. In Deutschland reicht es für das Bio-Siegel aus, auf den Einsatz von Pestiziden in der Lebensmittelproduktion zu verzichten. Bio-Siegel auf dem deutschen Markt verlangen von den Erzeugern zusätzliche qualitätssichernde Maßnahmen, wie den Verzicht auf leistungssteigernde Medikamente in der Fleischproduktion. Yormirzoev et al. (2020) fanden auf dem russischen Markt eine höhere ZB für Bio-Milch im Vergleich zu anorganischer Milch und keine Unterschiede zwischen den ZB-Werten für gekennzeichnete und nicht gekennzeichnete Bio-Milch.

Die Platzierung von Produkten am Point of Sale (POS) kann die Wahrnehmung dieser Produkte und ihren Wiedererkennungswert beeinflussen, was wiederum das Kaufverhalten beeinflusst. Lebensmittelqualität und -beschaffenheit sind wichtige Entscheidungskriterien. Beide Produkteigenschaften werden durch Attribute wie die Warenpräsentation am POS und die Art der Verpackung beeinflusst. Der Sicherheitsaspekt beim Verzehr verderblicher Produkte hat laut einer Studie über den sri-lankischen Markt (Ranasingha et al. 2019) einen starken Einfluss auf die ZB für lokale Früchte wie Trauben und Granatäpfel. Das Preisbewusstsein der Verbraucher stellt ein Hindernis für den Kauf von Bio-Lebensmitteln dar und trägt zur Verbreitung von Labels für konventionelle Lebensmittel bei (Olbrich et al. 2014). Diese Entwicklung erfordert eine eingehende Untersuchung der Bedeutung von Labels und POS im Kaufprozess und ihrer Auswirkungen auf Präferenzstrukturen.

Eine weitere Determinante der ZB für Bio-Produkte ist die Auswirkung der individuellen Einstellung zu Lebensmitteln auf die Verbraucherpräferenzen. Asante-Addo und Weible (2019) fanden in Ghana signifikante und heterogene Auswirkungen der Einstellung zu Lebensmitteln auf die ZB der Einwohner. Da nationale Bedingungen wie Lebensmittelgesundheit und Qualitätsstandards die individuellen Erwartungen und Präferenzen beeinflussen können, müssen die geografischen Auswirkungen dieser Determinante weiter erforscht werden.

Wir greifen die Ergebnisse von Paul und Rana (2012) auf, indem wir davon ausgehen, dass Verbraucher Bio-Lebensmittel kaufen, wenn der Gesamtnutzen die Kosten des Konsums von Bio-Lebensmittel überwiegt. Zur Untersuchung wird ein DCE angewendet, um die marginale ZB (MZB) für einzelne Bio-Attribute und spezifische Stufen zu ermitteln. Dadurch werden die von Asante-Addo und Weible (2019), Paul und Rana (2012), Bazzani et al. (2017), Marreiros (2021) sowie Meyerding (2018) verwendeten Modelle um die Variable Haushaltsmitglieder erweitert, um zu untersuchen, ob diese Variable die Nutzenfunktion im MNL-Modell beeinflusst. Das zusätzlich verwendete LC-Modell ermöglicht es, heterogene Präferenzstrukturen durch die Bildung von Klassen mit den dazugehörigen Zugehörigkeitsfunktionen aufzudecken und die Gültigkeit der Ergebnisse für den deutschen Markt zu überprüfen. Basierend auf diesen Zusammenhängen wurden die folgenden Hypothesen zu den Auswirkungen auf die Nutzenfunktion für Bio-Lebensmittel entwickelt:

- H1: Eine steigende Haushaltsgröße hat einen positiven Effekt auf die Nutzenfunktion.
- H2: Weibliche Konsumenten haben eine höhere Präferenz für Bio-Lebensmittel als männliche.
- H3: Jüngere Verbraucher haben eine höhere Präferenz für Bio-Lebensmittel.
- H4a: Ein steigendes Nettoeinkommen wirkt sich positiv auf die Nutzenfunktion aus.
- H4b: Steigende Durchschnittsausgaben haben einen positiven Effekt auf die Nutzenfunktion.
- H5: Lokale und regionale Produkte haben den höchsten positiven Einfluss auf die ZB.
- H6: Bio-Gütesiegel wirken sich positiv auf die ZB aus.
- H7: Der Verkaufsort hat einen Einfluss auf die ZB.
- H8: Unverpackte Produkte haben den stärksten Einfluss auf die ZB.
- H9: Die Marke des Herstellers hat die höchste ZB im Vergleich zu anderen Marken.

3 Forschungsmodell

Die soziodemographischen Determinanten des ZB für einzelne Attribute der untersuchten Produkte oder Dienstleistungen in hypothetischen Marktsituationen können geschätzt werden und der Nutzen der Verwendung eines Produkts kann in einzelne Attribute wie Qualität, Art der Inhaltsstoffe und im Produktionsprozess verwendete Faktoren zerlegt werden.

3.1 Erhebungsdesign und DCE-Aufbau

Der DCE ist in einen Fragebogen eingebettet, der Fragen zur aktuellen Lebenszufriedenheit, zur Einstellung zu Bio-Lebensmitteln und zu persönlichen Daten enthält. Die Stichprobe wurde von der Firma Omnipanel durchgeführt, und die Befragten wurden über den Grund für die Durchführung dieser Un-

tersuchung, die Verwendung ihrer Daten und darüber informiert, dass mit der Teilnahme an der Umfrage keine Risiken verbunden sind. Es wurde ein Datensatz von $N = 3.200$ mit einer probabilistischen Stichprobe von 400 und jeweils acht Auswahlätzen erstellt (Lizin et al. 2022; deBekker et al. 2015). Die Anzahl der Attribute und die entsprechenden Ebenen werden durch gegenläufige Faktoren bestimmt: Je mehr Attribute verwendet werden, desto höher ist der Informationsgehalt der Befragten. Allerdings steigt damit auch der zeitliche und inhaltliche Aufwand für die Befragten (DeShazo und Fermo 2002). Die definierten Attribute und Ebenen spiegeln die Erkenntnisse aus der in Abschnitt 2 beschriebenen Literatur wider und werden als relevant für die Verbraucher angenommen. Bei der Auswahl der Attribute wurden Studien wie die von Johnston et al. (2017) und DeShazo und Fermo (2002) berücksichtigt, um die Anzahl der Attribute auf die zentralen Eigenschaften von Bio-Lebensmitteln zu beschränken.

Dies geschieht, um Verzerrungen durch Informationsverzerrungen und Nichtbeachtung von Attributen zu reduzieren und die Konstruktvalidität zu erhöhen (Mariel 2021). Der Fragebogen und das Auswahl-design wurden im Rahmen einer Vorstudie von Studierenden in Bachelorstudiengängen entwickelt und gestaltet. Wir gehen davon aus, dass alle Attribute das individuelle nutzenmaximierende Verhalten eines Konsumenten bestimmen. Beim ersten Attribut "Herkunft" wird davon ausgegangen, dass die Konsumenten von Bio-Lebensmitteln und Getränken sensibel für Nachhaltigkeit sind. Diese Annahme erhöht den Nutzen dieses Attributs mit dem Grad der Lokalität, d. h. je kürzer die Wertschöpfungskette, desto höher der erwartete Nutzen. Bei "Verpackung" und "POS" werden Effekte der Verpackungsart und des Verkaufsortes auf den wahrgenommenen Nutzen angenommen. Entsprechend dem Maximierungsverhalten der Haushalte bestimmt das Vorhandensein eines "Gütesiegels" und dessen Qualitätsstandard die Wirkung auf die Nutzenfunktion. Bei "Marke" wird angenommen, dass der wahrgenommene Status einer Marke die Nutzenfunktion bestimmt. Das Attribut "Preis" ist in DCE Standard und wird für die Berechnung des MWTP für einzelne Attribute in Bezug auf ein vorgegebenes Standardproduktdesign benötigt.

Die Auswahl der Attribute basiert auf einer Vorstudie und wurde durch Studien zum Konsumentenverhalten zu Bio-Lebensmitteln ergänzt (Hempel/Hamm 2016). Jedes Attribut hat, je nach Ausprägung, drei bis sechs Stufen. Die Variation der Stufen für jedes Attribut (Tabelle 1) ermöglicht die Konstruktion einer hypothetischen Marktsituation und die Untersuchung der angegebenen Präferenzen der Verbraucher.

Tabelle 1: Attribute und Level

Attribute	Level
Herkunft	lokal; regional; national; EU-weit; weltweit
Verpackung	nicht verpackt oder selbstverpackt; Karton oder Papier; Glas; Kunststofffolie; Hartplastik
Verkaufsstelle	nach Kategorie / nur Bio-Produkte; separate Plätze für Bio-Lebensmittel; In-Shop-Bioladen
Label	ohne; nach EU-Standard; Deutsche Gütesiegel z. B. Bioland, Demeter, Naturland ¹
Marke	ohne Markenzeichen oder Marke unbekannt; Eigenmarke, z. B. „ja“, „Naturgut“; Herstellermarke, z. B. „Alnatura“; Premiummarke, z. B. „Bio Gourmet“
Maximaler Preisaufschlag	0 %; 25 %; 50 %; 75 %; 100 %

Quelle: eigene Tabelle

Zwei generische Alternativen und eine Nicht-Kauf-Option werden präsentiert, wenn die Befragten nicht bereit sind, eine der Optionen auf der Auswahlkarte zu wählen. Die Befragten wurden gebeten, jede der acht vorgelegten Auswahlkarten zu beantworten. Die Kombinationen der einzelnen Attribute und der dazugehörigen Stufen wurden zufällig ausgewählt. Jedes Attribut und jede Stufe repräsentiert die Merkmale von Bio-Produkten, die für die Wahl des Verbrauchers am POS von großer Bedeutung sind. Der Schwerpunkt liegt auf den in Deutschland ständig erhältlichen Bioprodukten: Milch, Fleisch, Äpfel, Kartoffeln und Wein. Die Analyse der Ergebnisse bezieht sich auf eine Standardvariante der ausgewählten Produkte aus konventioneller Produktion ohne zusätzliche Stufen. Das folgende Beispiel ist eine stochastische Kombination von sechs Attributen mit jeweils einer vordefinierten Stufe.

¹ Siehe z. B. Bio-Siegel: <https://www.oekolandbau.de/en/bio-siegel/>.

Tabelle 2: Beispiel für eine Wahlkarte

Attribut	Produkt A	Produkt B	Kein Kauf
Herkunft	regional	lokal	Ich wähle keine Option
Verpackung	Kunststofffolie	Unverpackt	
Verkaufsstelle (POS)	nach Kategorie / nur Bio-Produkte	separate Plätze für Bio-Lebensmittel	
Label	Keins	EU-Standard	
Marke	Herstellermarke	Premiummarke	
Preisaufschlag (%)	100	0	
Ich wähle	☐	☐	☐

Quelle: eigene Tabelle

Das Softwarepaket NGene (ChoiceMetrics 2012) wurde verwendet, um die Attribute und Merkmale der Produktalternativen für ein geeignetes statistisches Design anzuordnen. Es wurde ein effizientes fraktionales faktorielles Design erstellt, um Korrelationen zwischen den Attributsebenen auf der Grundlage von Erkenntnissen aus der Literatur auszuschließen (Bliemer/Collins 2016). Die Standardfehler der zu schätzenden Parameter wurden für ein multinomiales Logit-Modell minimiert (d-Effizienzkriterium) (Rose/Bliemer 2008). Die Prioritäten wurden auf der Grundlage von Erkenntnissen aus der Literatur gewählt (Mariel et al. 2021). Unplausible und dominante Kombinationen wurden ausgeschlossen.

3.2 Modellbeschreibung

Diskrete Auswahlexperimente beruhen auf hypothetischen Entscheidungen, die anhand einer befragten Stichprobe der Zielpopulation getroffen werden. Es wird angenommen, dass die Entscheidungen der Befragten den Nutzen maximieren, und jede Alternative i kann durch eine Nutzenfunktion U_i dargestellt werden. Zur Analyse der Daten wurde ein Zufallsnutzenmodell (RUM) verwendet (McFadden, 1974). In unserem theoretischen Modell der Kaufentscheidung gehen wir davon aus, dass der Gesamtnutzen eines Bio-Produkts die Summe des Nutzens der einzelnen Attribute und eines konstanten Nutzens ist. Der Nutzen eines Bio-Produkts lässt sich zum Beispiel wie folgt beschreiben:

$$U_x = \beta_n + \beta_1 orig + \beta_2 pack + \beta_3 pos + \beta_4 seal + \beta_5 bran + \epsilon_i. \tag{1}$$

Der Term U_x steht für den Nutzen, den ein Befragter aus dieser spezifischen Konfiguration des Bio-Produkts zieht, β_n steht für den Nutzenzuwachs oder -verlust eines Bio-Produkts in seiner Basiskonfi-

guration im Vergleich zu keinem Konsum, und die übrigen β -Parameter stehen für den aus den entsprechenden Attributen abgeleiteten Nutzen. ϵ_i ist der nicht-beobachtbare Teil von U_x , der den Fehlerterm darstellt. Für jede Option in der Auswahlmenge wählt der Befragte die Option, die den höchsten Nutzen bietet. Um die Zahl der Beobachtungen in unserer Studie zu erhöhen, wurde jeder Befragte achtmal gebeten, diese Wahl unter Verwendung verschiedener Attributwerte zu treffen. Die Anwendung der DCE basiert auf den methodischen Standards des RUM und verwendet das Discrete-Choice-Regressionsmodell (McFadden 1974) und das LC-Modell zur Schätzung des ZB und zur Untersuchung der Präferenzheterogenität (Truong/Hensher 2014; Rousseau 2015).

Der ZB ist der Geldbetrag, den die Verbraucher bereit sind, für ein Produkt zu zahlen, das die zugrundeliegende Eigenschaft aufweist, z. B. Lebensmittel aus der Region oder nach zertifizierten Standards hergestellte Bioprodukte.

4 Ergebnisse des Modells

Für die statistische Analyse haben wir die Software R-Studio verwendet. Mit Hilfe des Apollo-Pakets führten wir ein MNL- gefolgt von einem LC-Modell durch, um die Auswirkungen der einzelnen Attribute und der sozioökonomischen Variablen auf das Wahlverhalten zu schätzen und die MWTP für die einzelnen Attribute zu ermitteln. Tabelle 3 veranschaulicht die persönlichen Daten.

Tabelle 3: *Profil der Stichprobe*

Variable	Mittelwert	Std.-Abw.
Gesundheitsstatus: 1 = sehr gut, 2 = gut, 3 = zufriedenstellend, 4 = weniger gut, 5 = schlecht	2,44	0,99
Haushaltsmitglieder	2,2	1,2
Alter	49,8	15,5
Weiblich = 48.3 %, männlich = 51.7 %		
Wohnort: ländlich = 26 %, Kleinstadt = 36.3 %, Stadt = 37.7 %		
Durchschnittliche Ausgaben pro Monat für konventionelle Lebensmittel und Getränke in €	2.911	1.923.5
Durchschnittliche Ausgaben pro Monat für Bio-Lebensmittel und Getränke in €	682,2	824,5

Quelle: *eigene Tabelle*

4.1 Multi Nominal Logit und Latent Class Model

Tabelle 4 gibt einen Überblick über die durchschnittliche Auswirkung der einzelnen Attribute auf die Nutzenfunktion und das Konfidenzintervall. Der negative und signifikante Koeffizient der ASC der Opt-Out-Alternative impliziert, dass die Befragten im Durchschnitt dazu neigen, in Bio-Produkte zu investieren, wenn alles andere gleichbleibt. Die Koeffizienten für das Attribut Regionalität sind hoch signifikant und positiv. National und regional produzierte Waren werden bevorzugt. Die Koeffizienten für das Attribut Verpackung sind ebenfalls signifikant, mit Ausnahme des Koeffizienten für die Ebene Kunststofffolie. Karton/Papier und, in geringerem Maße, Glasverpackungen werden gegenüber anderen Verpackungsmaterialien bevorzugt. Vor allem für Folienverpackungen haben die Befragten kaum eine Präferenz. Die Präsentation am POS spielt bei der Kaufentscheidung keine große Rolle, denn die Koeffizienten sind alle signifikant, aber niedrig. Der Koeffizient für ein Gütesiegel nach EU-Norm ist signifikant und deutet auf eine hohe Präferenz für dieses Attributniveau hin. Beim Attribut Marke wird die Herstellermarke gegenüber den anderen Stufen deutlich bevorzugt. Dieser Wert ist hoch signifikant.

Tabelle 4: Ergebnisse der MNL Modelschätzung

Attribute und Level	Koeffizient	Std. Abw.	95%-Konfidenzintervall	
ASC_A	-0,1827***	0,0701	-0,1913	-0,1740
ASC_B	-0,184***	0,0701	-0,1927	-0,1754
Herkunft: lokal	0,916***	0,1177	0,8940	0,9379
Herkunft: regional	0,9478***	0,1277	0,9250	0,9706
Herkunft: national	0,4404***	0,1144	0,4182	0,4627
Herkunft: EU-weit	0,4809***	0,1070	0,4611	0,5008
Verpackung: unverpackt	0,8137***	0,1201	0,7928	0,8346
Verpackung: Karton/Papier	0,6383***	0,1263	0,6156	0,6609
Verpackung: Glas	0,5991***	0,1232	0,5755	0,6226
Verpackung: Kunststofffolie	0,1224	0,1006	0,1014	0,1434
POS: separate Plätze für Bio-Lebensmtl.	0,1285**	0,0704	0,1137	0,1432
POS: In-Shop-Bioladen	0,1086**	0,0633	0,0950	0,1222
Label: EU-Standard	0,1854***	0,0690	0,1725	0,1982
Label: Deutsch	0,2295***	0,0755	0,2157	0,2432

Marke: Eigenmarke	0,2635***	0,0757	0,2482	0,2787
Marke: Herstellermarke	0,3399***	0,0861	0,3246	0,3552
Marke: Premiummarke	0,2473***	0,0792	0,2307	0,2639
Maximaler Preisaufschlag	-0,0188***	0,0013	-0,0190	-0,0186
Soziodemografische Variablen	Koeffizient	Std. Abw.	95%-Konfidenzintervall	
ASC Haushaltsmitglied (H1)	0,3235***	0,0996	0,3126	0,3344
ASC Geschlecht (H2)	0,397**	0,1711	0,3770	0,4171
ASC Alter (H3)	-0,0115**	0,0057	-0,0122	-0,0108
ASC Nettoeink. pro Haushalt (H4a)	-0,0186	0,0262	-0,0217	-0,0155
ASC Durchschnittsausgaben für Bio-Lebensmittel und Getränke (H4b)	0,0006***	0,0001	0,0006	0,0006

Signifikanz (p-Werte): * = $p < 0,1$ ** = $p < 0,05$ *** = $p < 0,01$

Quelle: *eigene Tabelle*

Die negativen und signifikanten ASC-Werte zeigen, dass die Wahl der Standardalternative ohne nutzensteigernde Stufen zu einer Nutzenminderung führt. Abweichungen von dieser Standardalternative führen zu einer Erhöhung des Nutzens. Die positiven Werte der Koeffizienten für die einzelnen Stufen der Attribute Regionalität, Verpackung, Präsentation am POS, Siegel und Marke bestätigen diesen Nutzeneffekt, denn sie deuten darauf hin, dass eine Abweichung vom Standard über die jeweilige Stufe zu einer Nutzenerhöhung führt. Die vier Koeffizienten für das Attribut Regionalität sind alle auf dem 1 %-Niveau signifikant. Den größten Nutzenzuwachs haben lokal und regional produzierte Produkte im Vergleich zu Produkten, die global oder ohne räumlichen Bezug produziert werden. Die Merkmale nationale und EU-weite Produktion führen ebenfalls zu einem signifikanten Nutzenanstieg, der jedoch nur etwa halb so stark ausgeprägt ist. Im Vergleich zu Verpackungen aus Hartplastik, die hier als Standard gesetzt wurden, haben unverpackte Waren eindeutig den größten Nutzen. Im Vergleich zur Standardalternative ohne Biosiegel zeigt sich ein signifikanter positiver Effekt auf die Nutzenfunktion beim Kauf eines Produkts mit einem Biosiegel nach EU-Standard oder einem deutschen Siegel. In beiden Fällen sind die Ausprägungen ähnlich wie beim POS-Attribut. Auch die jeweiligen Marken der untersuchten Produkte haben einen signifikanten, aber nicht sehr ausgeprägten Nutzeneffekt gegenüber der Standardangabe keine oder unbekannte Marke. Die Verbraucher bevorzugen Herstellermarken wie „Alnatura“ gegenüber der Standardvariante. Etwas weniger stark ausgeprägt ist der positive Nutzeneffekt beim Kauf von Eigenmarken wie „ja“.

Die signifikanten Koeffizienten der sozioökonomischen Variablen weisen auf die heterogenen Auswirkungen dieser Werte auf die Nutzenfunktion hin. Haushaltsmitglieder und Geschlecht zeigen einen

signifikant positiven Nutzeneffekt. Für die binäre Variable Geschlecht bedeutet dies, dass Frauen Abweichungen von der Standardalternative signifikant höher einschätzen als Männer. Mit zunehmendem Alter nimmt die Präferenz für Abweichungen von der Standardalternative leicht ab. Die signifikante Variable durchschnittliche Ausgaben für Bio-Lebensmittel und Getränke hat wenig Einfluss auf die Verbraucherpräferenzen. Preiserhöhungen werden von den Verbrauchern negativ bewertet, was im Einklang mit der mikroökonomischen Nutzentheorie steht.

Die Hypothesen H1 bis H4b wurden anhand der Ergebnisse des MNL-Modells getestet. H1, H2 und H3 können bestätigt werden, da die erwarteten Assoziationen signifikant sind und einen niedrigen Effektivwert aufweisen. Die Annahme, dass weibliche Konsumenten höhere Präferenzen für Bio-Lebensmittel haben als Männer (H2), stimmt mit den Ergebnissen von Printezis et al. (2019) überein. Darüber hinaus bestätigten Medici et al. (2021) die erwartete negative Korrelation des ZB mit dem Alter (H3). Der positive Effekt des Einkommens (H4a) wird verworfen, da der Koeffizient nicht signifikant ist. Die Auswirkungen des Einkommens auf die ZB sind in der Literatur uneinheitlich. Boccaletti und Nardella (2000) zeigten einen signifikanten Effekt, während Medici et al. (2021) keinen signifikanten Effekt fanden. Die Annahme, dass eine Erhöhung der durchschnittlichen Ausgaben einen positiven Effekt auf die Nutzenfunktion für Bio-Lebensmittel (H4b) hat, wird akzeptiert, allerdings mit sehr geringem Effekt, was mit den Ergebnissen von Kabir et al. (2023) übereinstimmt.

Die Ergebnisse des LC-Modells bestätigen die Annahme der Heterogenität zwischen den gebildeten Klassen. Die Wahrscheinlichkeit, Mitglied der ersten Klasse zu sein, beträgt 32,1 % und die Wahrscheinlichkeit, dass eine Person zur zweiten Klasse gehört, 67,9 %. In Klasse eins sind nur für das Attribut Verpackung die Koeffizienten für zwei von drei Stufen und das Attribut Preis auf dem 1 %-Niveau signifikant. In der zweiten Klasse sind alle Einzelwerte auf dem 1 %-Niveau signifikant, mit Ausnahme einer Stufe des Attributs Siegel. Ähnlich wie beim MNL-Modell haben national und regional hergestellte Produkte den stärksten Nutzen. Auch in der EU hergestellte Produkte erhöhen den Nutzen signifikant. Unverpackte Produkte und Verpackungen aus Karton/Papier, Kunststoff oder Glas werden gegenüber dem Standard deutlich bevorzugt. Alle anderen Merkmalsausprägungen bewirken ebenfalls positive Nutzeneffekte, allerdings in deutlich geringerem Umfang. Lediglich die Herstellermarke hat einen etwas stärkeren Einfluss auf die Nutzenfunktion. Die Preiserhöhungen werden erwartungsgemäß für beide Klassen negativ bewertet. Mit der Klasse zwei als Referenz haben Männer eine höhere Wahrscheinlichkeit, Mitglied der Klasse zwei zu sein. Je größer die Anzahl der Haushaltsmitglieder ist, desto größer ist die Wahrscheinlichkeit, der Klasse zwei anzugehören. Je älter ein Befragter ist, desto größer ist die Wahrscheinlichkeit, dass er zur ersten Klasse gehört. Je höher die Bio-Ausgaben, desto größer ist die Wahrscheinlichkeit, zur zweiten Klasse zu gehören. In allen anderen Fällen hatten die individuellen Merkmale keinen signifikanten Einfluss auf die Klassenzugehörigkeit.

Sowohl die BIC- als auch die AIC-Werte des LC-Modells waren niedriger als die des MNL-Modells. Dies deutet auf eine bessere Anpassung des LC-Modells als des MNL-Modells hin.

4.2 Marginale Zahlungsbereitschaftswerte

MWTP-Werte drücken den Kompromiss zwischen einem Attribut und den Investitionskosten aus, die den individuellen Nutzen auf dem gleichen Niveau halten würden. Die MWTP-Werte wurden als prozentualer Anteil der Investitionskosten für eine Steigerung eines Attributs um eine Einheit ausgedrückt. Tabelle 5 zeigt die MWTP-Schätzungen auf der Grundlage der MNL- und LC-Modelle. Die EW-Werte lagen bei beiden Modellen innerhalb akzeptabler Grenzen; daher wurden sinnvolle Werte angenommen.

Erwartungsgemäß sind die KVA für die regionalen und deutschlandweiten Attributwerte im MNL-Modell am höchsten. Dieser zusätzliche KVA deckt sich mit den Ergebnissen von Meyerding (2018). Für national produzierte Produkte ist der KVA etwas höher. Diese Zuschläge entsprechen in etwa den aktuellen Preisunterschieden zu konventionellen Produkten im Lebensmittelhandel in der Schweiz (BLW 2023).

Tabelle 5: Marginale ZB-Schätzungen basierend auf dem MNL- und auf dem LC-Modell

	MNL		Klasse 1		Klasse 2	
Attribute	MZB	S. A.	Mittelwert	S. A.	Mittelwert	S. A.
Herkunft: lokal	48,7*** (47,3 – 50,1)	7,1	8,80*** (7,8 – 9,8)	4,90	91,50*** (88,3 – 94,7)	16,50
Herkunft: regional	50,4*** (48,9 – 51,9)	7,5	6,10 (4,5 – 7,7)	8,20	92,20*** (88,7 – 95,7)	17,70
Herkunft: national	23,4*** (22,2 – 24,6)	6,3	-5,30 (-6,5 – -4,1)	5,90	55,30*** (52,3 – 58,3)	15,10
Herkunft: EU-weit	25,6*** (24,4 – 26,8)	6,2	11,40*** (10,4 – 12,4)	5,30	15,50* (13,7 – 17,3)	9,30
Verpackung: unverpackt	43,3*** (42,0 – 44,6)	6,5	23,80*** (22,3 – 25,3)	7,40	66,80*** (64,4 – 69,2)	12,40
Verpackung: Karton/Papier	33,9*** (32,5 – 35,3)	7,0	12,20*** (11,0 – 13,4)	6,10	76,90*** (74,1 – 79,7)	14,10
Verpackung: Glas	31,9*** (30,6 – 33,2)	6,7	5,90 (5,0 – 6,8)	4,80	74,00*** (71,6 – 76,4)	12,10

Verpackung: Kunststoffolie	6,5 (5,5 – 7,5)	5,3	-7,80 (-9,4 – -6,2)	8,10	25,40*** (24,0 – 26,8)	6,90
POS: separate Plätze für Bio-Lebensmittel	6,8* (6,1 – 7,5)	3,8	-6,20 (-7,2 – -5,2)	4,90	24,70*** (23,5 – 25,9)	5,90
POS: In-Shop-Bioladen	5,8* (5,2 – 6,4)	3,3	-4,40 (-5,2 – -3,6)	4,00	17,00*** (16,1 – 17,9)	4,70
Label: EU-Standard	9,9*** (9,2 – 10,6)	3,7	-9,00*** (-10,1 – -7,9)	5,40	-7,90 (-9,2 – -6,6)	6,70
Label: Deutsch	12,2*** (11,4 – 13,0)	4,1	-10,20*** (-11,0 – -9,4)	4,30	3,70 (2,9 – 4,5)	4,30
Marke: Eigenmarke	14*** (13,2 – 14,8)	4,2	-10,60*** (-11,8 – -9,4)	5,90	-6,50 (-7,6 – -5,4)	5,80
Marke: Herstellermarke	18,1*** (17,2 – 19,0)	4,7	-8,80*** (-9,8 – -7,8)	5,30	12,80*** (11,7 – 13,9)	5,40
Marke: Premiummarke	13,2*** (12,4 – 14,0)	4,3	-12,00*** (-13,2 – -10,8)	5,90	22,30*** (21,1 – 23,5)	6,10

95%-Konfidenzintervalle in Klammern. * = $p < 0,1$ ** = $p < 0,05$ *** = $p < 0,01$

Quelle: eigene Tabelle

Empirische Untersuchungen zur Akzeptanz von Preisaufschlägen für Bio-Lebensmittel in Deutschland lagen nur für das Jahr 2017 vor. Die Ergebnisse zeigen, dass etwa 50 % der Verbraucher bereit sind, einen Aufschlag zwischen 5 % und 10 % zu zahlen (Statista 2023). Auffällig sind die deutlich niedrigeren MWTP für die anderen Stufen dieses Attributs. Darüber hinaus fand Meyerding (2018) Unterschiede von bis zu ca. 50 % zwischen ausgewählten lokalen und EU-weiten Bio-Lebensmitteln. Dieser Befund deckt sich mit den Ergebnissen der Meta-Regressionsanalyse von Printezis et al. (2019), die zeigen, dass es einen ZB für lokale Produkte mit Mehrwert im Vergleich zu lokalen Produkten gibt. Das Attribut Verpackung ist ebenfalls durch einen heterogenen ZB gekennzeichnet. Bou-Mitri et al. (2021) untersuchten die Verbraucherpräferenzen für das Design von Lebensmittelverpackungen im Libanon und fanden heraus, dass etwa 59 % bereit waren, mindestens 3 % mehr für eine bessere Verpackung zu zahlen. Der hohe MWTP für dieses Attribut in Deutschland im Vergleich zu dem MWTP in weniger entwickelten Ländern könnte durch die hohe Präferenz für Produkte mit wenig Verpackungsmaterial

erklärt werden (Marken/Wagenfeld 2020). Die beiden Stufen des POS-Attributs wurden gegenüber der Standardvariante bevorzugt. Die Wirksamkeit von Produktqualität und -sicherheit kann durch die Präsentation am POS positiv oder negativ beeinflusst werden (Ellison et al. 2016). Zusammen mit unseren eigenen Ergebnissen und denen von Meyerding (2018) kann davon ausgegangen werden, dass es eine ausgeprägte MWTP für das Attribut Siegel gibt, die mit zunehmender räumlicher Entfernung abnimmt. Eine Befragung von Verbrauchern in Norwegen ergab, dass rund 55 % der Befragten bereit waren, für ein Bio-Siegel einen höheren Produktpreis zu zahlen (Gustavsen/Hegnes 2020). Die verschiedenen Marken werden von den Verbrauchern ähnlich bewertet. Diese Werte sind niedriger als die MWTP für regionale und nationale Herkunft und für eine Art von Verpackung, aber höher als für die anderen Attribute.

Heterogene Präferenzstrukturen wurden im LC-Modell sichtbar. Die MWTP-Werte waren in Klasse eins durchweg signifikant niedriger als in Klasse zwei. Mit Hilfe der Zugehörigkeitsfunktion lässt sich diese Heterogenität für die sozioökonomischen Parameter Geschlecht, Alter, Lebensstandard und Ausgaben für Bio-Lebensmittel und Getränke erklären. In Klasse eins lagen die MWTP-Werte für das Attribut Regionalität im Vergleich zur Standardvariante im oberen Bereich des Spektrums. Der höchste MWTP liegt für die Attributstufen Karton/Papier und Glas für das Attribut Verpackung vor. Die Attribute Gütesiegel und Marke haben negative Auswirkungen auf den MWTP.

In der zweiten Klasse stellt sich die Situation anders dar. Die Mitglieder waren bereit, für Lebensmittel aus der eigenen Region 91,5 % des Standardpreises zusätzlich zu zahlen. Für in Deutschland produzierte Lebensmittel liegt dieser Wert bei 92,2 %. Auch in der EU hergestellte Produkte sind für die Befragten mit einem KVA von 55,3 % attraktiv. Global hergestellte Produkte verursachen einen deutlich geringeren Nutzeneffekt, wobei dieser Wert nur auf dem 5 %-Niveau signifikant ist. Der MWTP für das Attribut Verpackung war ebenfalls heterogen. Der MWTP für unverpackte oder verpackte Produkte in Glas oder Karton/Papier war relativ hoch. Nur bei Kunststofffolien ist der KVA deutlich niedriger. Der MWTP für getrennte Flächen und Bioläden im Laden ist vergleichsweise niedrig und weist nur eine geringe Differenz auf. Beim Attribut Marke ist der MWTP für die Ebene Herstellermarken und Premiummarke hoch signifikant und im unteren Bereich des gesamten Spektrums.

H5 kann bestätigt werden, da die ZB für die Attributsebenen lokal und regional in den Modellen MNL und LC signifikant größer sind als für die anderen Ebenen. Dies steht im Einklang mit den Ergebnissen von Meyerding (2018) und Kabir (2023). H6 kann nur für das MNL-Modell bestätigt werden, da es einen signifikanten MWTP für ein Gütesiegel auf EU-Ebene und für Deutschland gibt. Im Gegensatz dazu zeigt das LC-Modell heterogene Präferenzen. Die Mitglieder der ersten Klasse geben negative ZB für beide Siegel an, während die Werte der zweiten Klasse nicht signifikant sind. Die Ergebnisse des MNL-Modells stimmen mit denen von Bazzani et al. (2017) überein. Im LC-Modell wird jedoch deutlich, dass die Zugehörigkeitsfunktion einen starken Einfluss auf die Präferenz für Qualitätssiegel hat. H7 (POS) wird

im MNL-Modell bestätigt. Im LC-Modell waren die ZB-Werte nur in Klasse zwei signifikant. Das ZB für unverpackte Produkte ist im MNL-Modell am höchsten und für Mitglieder der Klasse eins im LC-Modell. Daher wurde H8 im LC-Modell bestätigt, jedoch nicht für die zweite Klasse. Unter der Annahme, dass die Konsumenten auch für nachhaltige Verpackungsformen sensibilisiert sind, ist der hohe ZB für unverpackte Waren verständlich. Die ZB für diese Attributstufe ist in Klasse zwei nicht die höchste, aber die ZB für Karton oder Papier und für Glas sind kaum höher. H9 konnte in beiden Modellen bestätigt werden. Wir gehen davon aus, dass dies durch die Ergebnisse von (Kabir et al. 2023) zur Rolle der Rückverfolgbarkeit erklärt werden kann, da die Herstellermarken auf dem deutschen Markt mit diesen Qualitätseigenschaften werben.

5 Diskussion und Schlussfolgerung

5.1 Ergebnisse

Erstens zeigen die Ergebnisse, dass die Unterschiede zwischen den sozioökonomischen Variablen der Stichprobe und denen der Gesamtbevölkerung Deutschlands recht gering sind und daher als weitgehend repräsentativ angesehen werden können. Zweitens wird deutlich, dass die meisten der untersuchten Attribute und Ebenen einen signifikanten Einfluss auf die Nutzenfunktion des Befragten haben. Von den sozioökonomischen Variablen im MNL-Modell haben nur die Haushaltsgröße und das Geschlecht einen signifikanten und relevanten Einfluss auf die Nutzenfunktion. Im LC-Modell ist die Situation differenzierter und einige sozioökonomische Variablen beeinflussen die Wahrscheinlichkeit einer Klassenzugehörigkeit. Drittens zeigen die Modelle eine heterogene Verteilung des MWTP zwischen den beiden Modellen und innerhalb des LC-Modells. Darüber hinaus haben Vriezen et al. (2023) fanden dies in ihrer Metaanalyse der ZB für rückverfolgbare Lebensmittelprodukte für über 80 % der untersuchten Studien. Mitglieder der zweiten Klasse reagieren deutlich sensibler auf Produkteigenschaften, die mit Nachhaltigkeit in Zusammenhang stehen. Der MWTP für solche Attributstufen ist deutlich höher als der der Klasse eins, aber niedriger als der des MNL-Modells. Diese Ambivalenz steht im Einklang mit den Ergebnissen früherer Studien. Beispielsweise stellte Meyerding (2018) geringe Präferenzen für Qualitätssiegel fest, und Gustavsen und Hegnes (2020) zeigen eine Präferenz für dieses Attribut bei mehr als der Hälfte der Verbraucher. Auch die Vertrauenswürdigkeit von Bio-Siegeln spielt für viele Verbraucher eine wichtige Rolle. Kabir et al. (2023) zeigten, dass die meisten Verbraucher in Bangladesch wenig über den Informationsgehalt und die Überprüfbarkeit nachhaltiger Lebensmittel wissen, es aber eine ZB für ein transparentes Etikett gibt. Loureiro (2007) kommt zu vergleichbaren Ergebnissen für Verbraucher in Nordamerika und Stein und de Lima (2022) ermitteln diese Präferenzstruktur für EU-Länder. Die deutlich höheren ZB-Werte für die anderen Attribute der zweiten Klasse verdeutlichen das Marktpotenzial für Produkte mit einem hohen Maß an Attributen, die am stärksten mit Nachhaltigkeit verbunden sind. Beispielsweise kann ein regional hergestelltes Premium-

Markenprodukt, in Papier oder Pappe verpackt und in einem Bioladen platziert, fast zum doppelten Preis eines Standardprodukts angeboten werden. Es ist wichtig, die Präferenzstrukturen der jeweiligen Zielgruppen und deren jeweilige Klassenstrukturen zu kennen. Die Vermarktung von Bio-Lebensmitteln mit diesen Attributstufen sollte gezielt an die Präferenzstruktur der Zielgruppe angepasst werden. Die Ergebnisse der ZB-Analyse im LC-Modell legen insbesondere nahe, Frauen und jüngere Konsumenten mit größeren Ausgabenanteilen für Bio-Produkte in der Zielgruppe anzusprechen. Diese Ergebnisse für den deutschen Bio-Markt stimmen mit den Erkenntnissen von Kabir et al. überein. (2023), die einen höheren ZB für transparente Etiketten junger Verbraucher in Bangladesch bestätigten. Die in Abschnitt 4.2 dargestellte Kumulierung der jeweils höchsten ZB für einzelne Attributstufen führt zu etwa einer Verdoppelung des Produktpreises im Vergleich zum gleichen konventionell hergestellten Produkt. Zum Vergleich: Der Preis für Bio-Frischmilch lag in Deutschland zwischen 2020 und 2021 etwa 45 % über dem Preis für konventionelle Milch. Bei Naturjoghurt beträgt dieser Unterschied etwa 63 % (AMI, 2022). Im Jahr 2022 sank der Unterschied in der Milchproduktion von etwa 25 % auf 5 % (BLE 2023; Statista 2023). Für die Vermarktung des hier untersuchten Bio-Lebensmittels ist es daher wichtig, die ZB auf die einzelnen Attributstufen und die Präferenzstruktur der Zielgruppe auszurichten. Diese Maßnahmen sind sinnvoll, um dem permanenten Anstieg der Produktionskosten entgegenzuwirken. Anbieter von Bio-Lebensmitteln profitieren von den abnehmenden Preisunterschieden zu konventionellen Produkten sowie dem relativ stärkeren Anstieg der Marktpreise für vergleichbare konventionelle Produkte seit 2021 (BOELW 2023a). Darüber hinaus ist der Bio-Markt robuster, da er nicht auf knappe Mineraldünger mit volatilen Preisen angewiesen ist. Darüber hinaus sind Lieferketten in der Regel kürzer und weniger anfällig (BOELW 2023b). Für die weitere Analyse soziodemografischer Determinanten und regionaler Einflussfaktoren besteht auf der Grundlage von Vorkenntnissen Forschungsbedarf. Beispielsweise können Variablen im Zusammenhang mit den Lebensbedingungen, der Einstellung zu Ernährung und -politik sowie dem Wissensstand über Ernährungsaspekte die ZB für Bio-Attribute beeinflussen. Länderspezifische Präferenzbeschränkungen (Ranasingha et al. 2019; Paul/Rana 2012; Asante-Addo/Weible 2019) erfordern eine tiefere und umfassendere Analyse, um systematische Zusammenhänge zu untersuchen.

5.2 Einschränkungen

Erstens wird in der Umfrage bewusst eine hypothetische Marktsituation konstruiert, in der die Befragten lediglich ihre Konsumwahl äußern. Ein tatsächliches Kaufverhalten wurde nicht beobachtet. Dies kann zu einem hypothetischen Bias (HB) führen, durch den die Ergebnisse der DCE über- oder unterschätzt werden können. Hensher (2010) zeigte, dass unvermeidbare Unterschiede zwischen hypothetischen Experimenten und nicht-experimentellen Beweisen durch das entsprechende Design des DCE reduziert werden können. Eine Metaanalyse von Haghani et al. (2021) zeigten in zahlreichen Studien das Vorhandensein des HB, aber auch Unterschiede zwischen den Fachgebieten. Offensichtlich verrin-

gert die Bedeutung der Wahlalternativen die Diskrepanz zwischen angegebenen und offenbarten Präferenzen (Ha-ghani et al. 2021). Um den HB zu reduzieren, haben wir die Attribute und Niveaus in einer Vorstudie mit Studierenden ausgewählt und die Ergebnisse mit denen vergleichbarer Studien verglichen. Zweitens könnte durch den Vergleich der sozioökonomischen Daten der Stichprobe mit denen der Bevölkerung das Problem der systematischen Verzerrung aufgrund mangelnder Repräsentativität verringert werden. Drittens entstehen Status-Quo-Effekte, wenn Menschen es vorziehen, dass alles beim Alten bleibt, was bei einem Umfragedesign mit zwei Alternativen und einer Status-Quo-Option der Fall sein kann (Oehlmann et al. 2017). Wir haben diese Verzerrung reduziert, indem wir Alternativen abgegrenzt und die Anzahl der Ebenen minimiert haben. Das Problem der Nichtbeachtung von Attributen wurde durch die Auswahl von Attributen und Ebenen reduziert, indem das Auswahldesign für die Befragten verständlich und so unkompliziert wie möglich gestaltet wurde (Alemu et al. 2013). Eine weitere Einschränkung stellt der Einsatz einer Online-Umfrage dar, da hier nicht hinreichend sichergestellt werden kann, ob die Befragten die Gesamtsituation hinreichend verstanden haben. Die Literaturanalyse lieferte erste Hinweise auf die Validität und Zuverlässigkeit der Ergebnisse. Weitere Studien sind erforderlich, um die Effizienz des Auswahldesigns zu steigern und die Plausibilität der Ergebnisse zu erweitern. Die Ergebnisse des DCE bestätigen die früheren Studien und konkretisieren Verbraucherpräferenzen für die relevanten Produkteigenschaften von Bio-Lebensmitteln in Deutschland. Trotz dieser Einschränkungen liefern die Ergebnisse Einblicke in die Zusammenhänge zwischen ZB und sozioökonomischen Variablen und können von Einzelhändlern und Produzenten zur Verbesserung von Marketingkonzepten genutzt werden.

Ethische Konformität: Die Autoren bestätigen, dass bei der Durchführung der Studie die entsprechenden Protokolle zum Schutz der Rechte und der Privatsphäre aller Teilnehmer angewandt wurden.

Literaturverzeichnis

- Absatzwirtschaft (2022): Bei Bio-Lebensmitteln variieren die Preise je nach Bio-Supermarkt um 45 Prozent. <https://www.absatzwirtschaft.de/bei-bio-lebensmittel-schwanken-die-preise-um-45-prozent-je-nach-bio-supermarkt-98725/>. Download 02.08.2022.
- Alemu, M.H.; Mørkbak, M.R.; Olsen, S.B., & Jensen, C. L. (2013): Attending to the reasons for attribute non-attendance in choice experiments. *Environmental and Resource Economics*, 54(3), 333–359.
- AMI (2022): Consumer Price Index. <https://milchland.de/bio-vs-konventionell-verbraucherpreisvergleich-anhand-ausgewaehlter-milchprodukte/>. Download 01.03.2023.
- Angeliki, N. M.; Olsen, S. B. & Tsagarakis, K. P. (2016): Towards a common standard – A reporting checklist for web-based stated preference valuation surveys and a critique for mode surveys. *Journal of Choice Modelling*, 18, 18–50.
- Asante-Addo, C. & Weible, D. (2019): Is there hope for domestically produced poultry meat? A choice experiment of consumers in Ghana. *Agribusiness*, 36, 281–298.

- Balderjahn, I.; Peyer, M.; Seegebarth, B.; Wiedmann, K.-P. & Weber, A. (2018): The many faces of sustainability-conscious consumers: A category-independent typology. *Journal of Business Research*, 91, 83–93. <https://doi.org/10.1016/j.jbusres.2018.05.022>.
- Bazzani, C.; Caputo, V.; Nayga, R. & Canavari, M. (2017): Revisiting consumers' valuation for local versus organic food using a non-hypothetical choice experiment: Does personality matter? *Food Quality and Preference*, 62, 144–154.
- Blackwell, R. D.; Miniard, P. W. & Engel, J. F. (2008): *Consumer behaviour*, 10th ed. Forth-Worth. TX: The Dryden Press.
- BLE (Federal Office for Agriculture and Food). (2023): Preis ökologisch erzeugter Kuhmilch in Deutschland von Dezember 2020 bis Dezember 2022. In Statista. <https://de.statista.com/statistik/daten/studie/380536/umfrage/preis-oekologisch-erzeugter-kuhmilch-in-deutschland/?locale=de>. Download 01.03.2023.
- BLW (Bundesamt für Landwirtschaft) (2023): Fachbereich Marktanalysen, Nielsen, Switzerland. In: Statista. <https://de.statista.com/statistik/daten/studie/474974/umfrage/preisunterschied-zwischen-bio-und-konventionellen-lebensmitteln-in-der-schweiz/?locale=de>. Download 15.02.2023.
- BMEL (Federal Ministry of Food and Agriculture) (2022): *Eco-Barometer 2021*. Berlin.
- BMEL (Federal Ministry of Food and Agriculture) (2023): *Eco-Barometer 2022*. Berlin.
- Boccaletti, S.; Nardella, M. (2000): Consumer willingness to pay for pesticide-free fresh fruit and vegetables in Italy. *The International Food and Agribusiness Management Review*, 3(3): 297–310. [https://doi.org/10.1016/S1096-7508\(01\)00049-0](https://doi.org/10.1016/S1096-7508(01)00049-0).
- BOELW (2022): Share of organic food in food sales in Germany from 2010 to 2021 (Graph). In: Statista. <https://de.statista.com/statistik/daten/studie/360581/umfrage/marktanteil-von-biolebensmitteln-in-deutschland/>. Download 15.02.2023.
- BOELW (2023a): Price development of selected conventionally and organically produced milk products in Germany by sales line in autumn 2022. In: Statista. <https://de.statista.com/statistik/daten/studie/1366158/umfrage/preisanstieg-konventionell-oekologisch-milchprodukte-vertriebslinie/?locale=de>. Download 01.03.2023.
- BOELW (2023b): Organic is the answer to global crises. Unbroken consumer loyalty. Politicians must slow down on the 30% organic farming target (German press release). <https://www.boelw.de/presse/meldungen/artikel/boelw-bio-ist-die-antwort-auf-globale-krisen-verbrauchertreue-ungebrochen-politik-muss-beim-30-oeko-landbau-ziel-von-der-bremse-gehen/>. Download 01.03.2023.
- Bou-Mitri, C.; Abdessater, M.; Zgheib, H. & Akiki, Z. (2021): Food packaging design and consumer perception of the product quality, safety, healthiness and preference. *Nutrition & Food Science* 51(1), 71–86.
- Chip (2022): Wo-gibt-es-das-beste-guenstigste-Bio-Angebot_174764979.html. <https://www.chip.de/news/Aldi-Lidl-Edeka-und-Co>. Download 02.08.2022.
- De Becker-Grob, E. W.; Donkers, B.; Jonker, M. F. & Stolk, E. A. (2015): Sample Size Requirements for Discrete-Choice Experiments in Healthcare: a Practical Guide, *Patient*, 8, 373–384.

- DeShazo, J. R. & Fermo, G. (2002): Designing choice sets for stated preference methods: The effects of complexity on choice consistency. *Journal of Environmental Economics and Management*, 44(1), 123–143.
- Ding, Y.; Veeman, M. M. & Adamowicz, W. L. (2012): The influence of attribute cutoffs on consumers' choices of a functional food. *European Review of Agricultural Economics*, 39(5), 745–769.
- Ellison, B.; Bernard, J. C.; Paukett, M. & Toensmeyer, U. C. (2016): The influence of retail outlet and FSMA information on consumer perceptions of and willingness to pay for organic grape tomatoes. *Journal of Economic Psychology*, 55, 109–119.
- European Union (2021): Average Age of Residents in the Member Countries in the Year 2021, <https://de.statista.com/statistik/daten/studie/248994/umfrage/durchschnittsalter-der-bevoelkerung-in-den-eu-laendern/>.
- Gustavsen, G. W.; Hegnes, A. W. (2020): Individuals' personality and consumption of organic food. *Journal of Cleaner Production*, 245, 118772.
- Haghani, M.; Bliemer, M. C. J.; Rose, J. M.; Oppewal, H. & Lancsar, E. (2021): Hypothetical bias in stated choice experiments: Part I. Integrative synthesis of empirical evidence and conceptualisation of external validity. Cornell University, Econometrics. Elsevier. <https://doi.org/10.1016/j.jocm.2021.100309>.
- Harold, J.; Bertsch, V. & Fell, H. (2019): Consumer preferences for end-use specific curtailable electricity contracts on household appliances during peak load hours. ESRI Working Paper No. 632.
- Heckert, N. A., et al. (2012): Nist/Sematech e-Handbook of Statistical Methods <http://www.itl.nist.gov/div898/handbook/>. Download 28.07.2022.
- Hempel, C.; Hamm, U. (2016): Local and/or organic: a study on consumer preferences for organic food and food from different origins. *International Journal of Consumer Studies*, 40(6), 732–741. <https://doi.org/10.1111/ijcs.12288>.
- Hensher, D. A. (2010): Hypothetical Bias, Choice Experiments and Willingness to Pay. *Transportation Research Part B Methodological*, 44(6) 735–752. DOI:10.1016/j.trb.2009.12.012.
- Hensher, D. A., Rose, J. M. & Green. W. H. (2006): *Applied Choice Analysis. A Primer*.
- Hossain, K. Z.; Xue, J. & Rabbany, M. G. (2021): Consumers' willingness to pay for GLOBALG.A.P. Certified Chicken: empirical evidence from a consumer survey in Bangladesh. *Food Control* 130 (March), 108397. doi: 10.1016/j.foodcont.2021.108397
- Howard, K. & Salkeld, G. (2009): Does Attribute Framing in Discrete Choice Experiments Influence Willingness to Pay? Results from a Discrete Choice Experiment in Screening for Colorectal Cancer. *Value in Health*, 12(2), 354–362.
- Hüttel, A. & Balderjahn, I. (2021): The coronavirus pandemic: A window of opportunity for sustainable consumption or a time of turning away? *Journal of Consumer Affairs*, 56, 68–96. <https://doi.org/10.1111/joca.12419>.
- Johnston, R. J.; Boyle, K. J.; Adamowicz, W.; Bennett, J.; Brouwer, R.; Cameron, T. A.; Hanemann, W. M.; Hanley, N.; Ryan, M.; Scarpa, R. & Tourangeau, R. (2017): Contemporary guidance for stated preference studies. *Journal of the Association of Environmental and Resource Economists*, 4(2), 319–405.

- Kabir, A. N. M. F.; Alam, M. J.; Begum, I. A.; McKenzie, A. M. (2023): Consumers' interest and willingness to pay for traceable vegetables- An empirical evidence from Bangladesh. *Future Foods* 7, 100214.
- Lancaster K. J. (1966): A new approach to consumer theory. *Journal of Political Economy* 74 (132), 132–157. DOI: 10.1086/259131.
- Lizin, S.; Rousseau, S.; Kessels, R.; Meulders, M.; Pepermans, G.; Speelman, S.; Vandebroek, M.; Van Den Broeck, G.; Van Loo, E. J. & Verbeke, W. (2022): The state of the art of discrete choice experiments in food research. *Food Quality and Preference*, 102, 104678.
- Loureiro, M. L.; Umberger, W. J. (2007): A choice experiment model for beef: what us consumer responses tell us about relative preferences for food safety, country-of-origin labeling and traceability. *Food Policy* 32 (4), 496–514. doi: 10.1016/j.foodpol.2006.11.006
- Mariel, P.; Hoyos, D.; Meyerhoff, J.; Czajkowski, M.; Dekker, T.; Glenk, K.; Bredahl Jacobson, J.; Liebe, U.; Olsen, S. B.; Sagebiel, J.; & Thiene, M. (2021): Environmental valuation with discrete choice experiments: Guidance on design, implementation and data analysis (129p). Springer Nature.
- Marken, G.; Wagenfeld, N. with the collaboration of Dangel, V., Hettich, E., Kitzberger, M., Nestele, J. & Rubik, F. (2020): Einkaufsgewohnheiten und Einflussfaktoren beim Unverpackt-Einkauf. Report of the Research Project Innoredux.
- Marreiros, C. G.; Dionísio, A. & Lucas, M. R. (2021): Does country matter in urban organic food products consumption? *International Journal of Consumer Studies*, 45. 1-13. DOI: 10.1111/ijcs.12599.
- Medici, M.; Di Noto, L.; Canavari, M. (2021): Role of sustainability attributes and price in determining consumers' fruit perceived value. *International Journal on Food System Dynamics*, 12 (4), 314–326. doi: 10.18461/ijfsd.v12i4.93
- Meyerding, S. (2018): Consumer Preferences for Organic Labels in Germany Using the Example of Apples – Combining Choice-Based Conjoint Analysis and Eye-Tracking Measurements. *Journal of Cleaner Production*, 181, 772-783. DOI: 10.1016/j.jclepro.2018.01.235.
- Oehlmann, M.; Meyerhoff, J.; Mariel, P. & Weller, P. (2017): Uncovering context-induced status quo effects in choice experiments. *Journal of Environmental Economics and Management*, 81, 59–73.
- Olbrich, R.; Hundt, M. & Grewe, G. (2014): Willingness to Pay in Food Retailing – An Empirical Study of Consumer Behaviour in the Context of the Proliferation of Organic Products. In: T. Foscht, D. Mor-schett, T. Rudolph, P. Schneditz, H. Schramm-Klein & B. Swoboda (2014): *European Retail Research* (pp. 67–101). Wiesbaden: Springer.
- Paul, J. & Rana, J. (2012): Consumer behavior and purchase intention for organic food. *Journal of Consumer Marketing*, 29(6), 412–422. DOI 10.1108/07363761211259223.
- Printezis, I.; Grebitus, C.; & Hirsch, S. (2019): The price is right!? A meta-regression analysis on willingness to pay for local food. *PLoS ONE*, 14(5). e0215847.
- Profeta, A.; Baune, M.-C.; Smetana, S.; Broucke, K.; van Royen, G.; Weiss, J.; Hieke, S.; Heinz, V.; Terjung, N. (2021): Consumer preferences for meat blended with plant proteins – Empirical findings from Belgium. *Future Foods* 4, 100088.
- Ranasingha, R. G. S. M.; Edirisinghe, J. C. & Ratnayake, K. (2019): Willingness to Pay for Fruit Attributes: A Conjoint Analysis. *The Journal of Agricultural Sciences – Sri Lanka*, 14(2), 102–110. <http://dx.doi.org/10.4038/jas.v14i2.8512>.

- Rose J. M. & Bliemer M. C. J. (2013): Sample size requirements for stated choice experiments. *Transportation*, 40, 1021–1041.
- Sarma, P. K. (2020): Investigating consumers' preference on fresh vegetables in Bangladesh: best-worst scaling approach. *Journal of Critical Reviews* 7 (11), 1821–1832. doi: 10.19080/AR-TOAJ.2020.24.556254.
- Sohn, S.; Seegebarth, B. & Woisetschlager, D. M. (2022): The same only different? How a pandemic shapes consumer organic food purchasing. *Journal of Consumer Behaviour*, 21(5), 1–14. DOI: 10.1002/cb.2060.
- Statista (2021a): Population in Germany by gender and generations 2021, <https://de.statista.com/statistik/daten/studie/1131309/umfrage/bevoelkerungsanteile-nach-geschlecht-und-generation/>. Download 02.08.2022.
- Statista (2021b): Population structure and development in rural* and urban areas in East and West Germany up to 2018, <https://de.statista.com/statistik/daten/studie/1186462/umfrage/bevoelkerungsentwicklung-land-stadt-ost-west/>. Download 02.08.2022.
- Statista (2021c): Women and men in Germany by highest school leaving certificate in comparison with the population in 2021, <https://de.statista.com/statistik/daten/studie/290386/umfrage/umfrage-in-deutschland-zu-frauen-und-maennern-nach-bildungsabschluss/>. Download 02.08.2022.
- Statista (2022a): Share of bio-food at the food revenue in Germany 2011 to 2022. <https://de.statista.com/statistik/daten/studie/360581/umfrage/marktanteil-von-biolebensmitteln-in-deutschland/?locale=de>. Download 15.02.2023.
- Statista (2022b): Differences in revenues with bio-food in Germany 2000 to 2022. <https://de.statista.com/statistik/daten/studie/4109/umfrage/bio-lebensmittel-umsatz-zeitreihe/?locale=de>. Download 28.02.2023.
- Statista (2023): <https://de.statista.com/infografik/24615/preisaufschlaege-fuer-bio-lebensmittel-in-deutschland/>. Download 15.02.2023.
- Statista (2023): <https://de.statista.com/statistik/daten/studie/172772/umfrage/akzeptanz-gegenueber-preisaufschlag-bei-bioprodukten-seit-2008/?locale=de>. Download 15.02.2023.
- Statista (2023): *Ökologische Lebensmittelwirtschaft – Branchenreport 2023*, p. 26. Download 12.05.2023.
- Statistik-Portal der Bundesländer und des Bundes (2021): <https://www.statistikportal.de/de/bevoelkerung/haushalte>. Download 28.10.2022.
- Stein, A. J.; de Lima, M. (2022): Sustainable food labelling: considerations for policy-makers. *Review of Agricultural, Food and Environmental Studies* 103, 143–160. <https://doi.org/10.1007/s41130-021-00156-w>.
- Train, K. (2002): *Discrete Choice Methods with Simulation*. National Economic Research Associates. Cambridge University Press.
- Truong, T. P. & Hensher, D. A. (2014): Linking discrete choice to continuous demand in a spatial computable general equilibrium model. *Journal of Choice Modelling* 12, 21–46.

- Varian, H. (2010): *Intermediate Microeconomics. A Modern Approach.* (8th ed.) W. W. Norton & Company.
- Vriezen, M.; Plishka, R.; Cranfield, J. (2022): Consumer Willingness to Pay for Traceable Food Products: a Scoping Review. *British Food Journal*. doi: 10.1108/BFJ-01-2022-0085.
- Wu, L.; Xu, L.; Zhu, D.; Wang, X. (2012): Factors Affecting Consumer Willingness to Pay for Certified Traceable Food in Jiangsu Province of China. *Canadian Journal of Agricultural Economics* 60 (3), 317–333. doi: 10.1111/j.1744-7976.2011.01236.x.

Autoren

Prof. Dr. Jens **Gericke** absolvierte das Studium der Wirtschaftswissenschaften mit dem Schwerpunkt Produktion/Logistik an den Universitäten Essen und Dortmund. Danach arbeitete er als Diplom Kaufmann mehrere Jahre im Logistikbereich eines süddeutschen Automobilherstellers. Anschließend erfolgte ein Wechsel in die Beratungsbranche. Nach einer langjährigen Beschäftigung als Unternehmensberater in Projekten unterschiedlichster Branchen im Produktions- und Logistikbereich erfolgte durch die Rückkehr an die Universität Dortmund der Einstieg in die Forschung und Lehre. Dort promovierte er zum Thema Outsourcing in der Logistik. Seit März 2005 ist er an der International School of Management (ISM) in Dortmund beschäftigt und lehrt dort als Professor in den Themenbereiche Unternehmensführung, Kosten- und Leistungsrechnung, Produktion und Logistik. Neben der Lehrtätigkeit gehört zu seinen Aufgabenbereichen u.a. die Entwicklung und Leitung verschiedener Studiengänge der ISM. Weiter führt er regelmäßig Workshops und Projekte mit mittelständischen Unternehmen durch.



Prof. Dr. Audrey **Mehn** ist Professorin für Marketing & Retail und Vizepräsidentin für Lehre an der ISM (University of Applied Sciences). Französischstämmig, studierte und erlangte Audrey Mehn ihren Doktorgrad an der Universität Siegen (Deutschland). Ihre Forschungsschwerpunkte umfassen den Wandel im Konsumentenverhalten, insbesondere in den Bereichen Mode, Kosmetik und Luxus.



Prof. Dr. Kai **Rommel** promovierte im Fach Agrarökonomie an der Landwirtschaftlich-Gärtnerischen Fakultät der Humboldt-Universität zu Berlin. Zwischen 2000 und 2007 habilitierte er sich am Lehrstuhl für Volkswirtschaftslehre, insbesondere Umweltökonomie an der Brandenburgischen Technischen Universität Cottbus in den Fächern Volkswirtschaftslehre, Energiewirtschaft und Umweltökonomie. Anschließend arbeitete er als Energiewirtschaftsstrategie bei der Energie Baden-Württemberg AG. Seit März 2010 ist er Professor und seit 2014 Vizepräsident für Forschung an der International School of Management in Dortmund. Er leitet Drittmittelprojekte im Bereich der Kommunalfinanzen und publiziert regelmäßig in Fachzeitschriften.



Westermann, Arne; Böckenholt, Rebecca; Ghorbani, Mijka; Messedat, Jons

Interior Architectural Branding – The Implementation of the Brand Identity at the Point of Sale by Means of Interior Design

Abstract

Today it becomes increasingly difficult to differentiate brands from competitors. A strong brand identity is a prerequisite for achieving this differentiation.

Existing research concerning the Point of Sale has focused primarily on the influence of store design on sales and consumer buying decisions, but not on its influence on brand perception. To fill this gap, we developed an explorative study design focusing on one exemplary brand.

We first examined how the brand personality is expressed in stores, second, we analysed the influence of the store design on the consumer perception of the brand with a representative online survey.

Our study shows that interior design elements are well suited to express the brand personality.

Summary Statement of Contribution

Existing research concerning the Point of Sale has focused on the influence of store design on sales and consumer buying decisions. The importance of the PoS for brand management and its influence on brand perception has not been examined so far. The study presented in this article fills this gap examining what influence the interior design has on brand perception. This research shows a strong influence on consumer's perception of brand values and attributes.

Keywords:

Brand identity, brand image, brand management, architectural branding, interior store design

1 Introduction

In an increasingly intense competitive environment, it is more and more difficult for companies to differentiate their brands from their competitors. A well thought-out brand concept and efficient brand management are the prerequisites for achieving this differentiation and for long-term success. (Burmam et al. 2018)

The concept of identity-oriented brand management (Meffert et al. 2005) is the guiding principle of brand management for many German brands and retailers: Brands need a strong identity and personality to be successful. The development of such a clear brand identity, which is based on the brand's self-concept and vision, is the fundamental task of brand management. The competencies, services and values as well as the brand appearance are derived from this self-concept. Already here, companies often face an obstacle, because often there is a lack of know-how and competence in this area.

But even when this task has been mastered, it is particularly important for brands to make this newly defined identity known to the target groups, especially when it comes to mass markets.

As a result, the second fundamental task of modern brand management is to communicate this identity with the help of all instruments of the marketing mix. The central goal is that the brand perception, i.e. the brand image, is as congruent as possible with the defined identity. Moreover, the brand must have a clear positioning that differentiates it from the competition. (Burmam et al. 2018)

The point of sale (PoS), i.e. the shop in the offline world, plays an outstanding role as a central point of contact with consumers. In addition to the behaviour of the sales staff, the ambience in these stores is decisive for the consumer perception (Müller 2012; Steiner 2011).

It is determined by the illumination, the scents used and by the music played – and also by the interior design, i.e. the way in which the walls, the ceiling and the floor are designed, and which decorative elements are used. However, the influence of these aspects on the perception of the brand identity, its values and attributes have not been in the centre of empirical research so far. This paper fills this gap introducing the results of an online survey representative for the German population between 18 and 64. In this survey, the impact of the interior design on the perception of one brand (the Dutch cosmetical company Rituals) has been examined using an experimental survey design. Furthermore, the empirical results concerning Rituals will be discussed with regard to this category of companies working out the central aspects with relevance for company practice.

2 Theoretical Basis

2.1 Identity-based Brand Management

Most theoretical considerations on brand management assume that brands need a strong identity with a clear and focused personality that distinguishes them from other brands (Aaker 1996; Burmann et al. 2018). The underlying basic consideration is that brands are perceived by consumers in a similar way that other people are. This is based on the idea of animism: people tend to attribute human characteristics to things, i.e. to humanise them. This also applies to brands. Jennifer Aaker first developed a complete theoretical approach in the 1990s (Aaker 1997). Starting with personality research in psychology, she derived five basic types of character traits for brands: These can therefore be characterised by sincerity, sophistication, robustness, competence or excitement, whereby brand personalities can also consist of combinations of these basic types (Aaker 1997).

At the same time, the concept of identity-based brand management was developed in German-speaking countries, which has some parallels to Aaker's core concept (Burmann et al. 2018; Esch 2018). The essential starting point here is the brand identity. It is understood as a determining factor, that makes the brand authentic and sustainably differentiates it from competitors (Burmann et al. 2018). Out of the various brand identity models (Aaker 1996; Kapferer 1997; Esch 2018; Burmann et al. 2018), the identity-based branding approach of Burmann et al. (2018) is most widespread in the German-speaking world and is also frequently applied (explicitly or at least immanently) in practice. Therefore, it forms the main theoretical basis for this paper.

In this context a brand is defined as a bundle of functional and non-functional benefits whose design differentiates itself sustainably from competing offers (brands) in the perception of the target groups of the brand (Burmann et al. 2017). The identity of the brand is determined by the vision, history, competencies, values and personality of a brand. Whereas the vision addresses the core idea of a brand and its future perspectives, the brand history describes where a brand comes from and what it has achieved in the past. Both, future and tradition, have influence on the competencies describing the core qualifications of a brand and on its values and personality. The values describe what a brand believes in, whereas the personality characterises the appearance of the brand.

In this context, it is particularly important to point out the importance of brand values and brand personality, as they not only represent a guideline for brand behaviour, but also decisively determine the perception of the philosophy of a brand (Burmann et al. 2018). They are therefore a crucial part of the conception of brand identity, and the external perception of the personality and of how the brand values are realised is a crucial factor for the evaluation of brand's credibility and its image. Therefore, the successful communication of the brand personality and the fulfilment of brand value promises is a basic prerequisite for the authenticity of a brand and its success (Burmann et al. 2018).

The major task of brand management is conceptualising, designing and managing the brand identity, whereby the brand's origin represents the foundation of the construct. This identity is then communicated to external target groups through brand touch points (Burmman et al. 2015). The aim is to harmonise the perception of the target group, i.e. the image, with the identity. The brand image can therefore be understood as a downstream reaction of the external target groups to brand management activities: Identity and image are thus self-image and external image of the brand (Burmman et al. 2015).

The basic prerequisite for the formation of an image is brand awareness consisting of brand recognition and brand perception. Since the formation of a brand image and the linking of associations are not possible without brand awareness, it is not defined as one of the actual components of the brand image, but as a necessary precondition for its development (Burmman et al. 2015).

The brand image itself can in turn be subdivided into the perceived brand attributes and benefits of the brand by the consumers. The (subjectively perceived) brand attributes describe the characteristics of a brand from the point of view of the target group. Here, on the one hand, knowledge about a brand is formed, e.g. about its history and the offered products. On the other hand, attributes are attached to the brand, which reflect the competencies (e.g. high quality) but also the appearance of a brand (e.g. sporty, young).

The extent to which the needs of consumers are satisfied by the brand depends primarily on the perceived brand benefits. These can be divided into a functional and a non-functional dimension (Burmman et al. 2015). The functional brand benefit describes the ability of the brand to solve concrete problems or to fulfil physical needs of the target groups, e.g. whether clothing of a certain brand is perceived as durable, pleasant to wear and warming. The non-functional use of a brand, on the other hand, is characterised on a psycho-social level: consumers express their (idealised) self-identity by buying or using brands. Thus, brands play a role in self-realisation. In addition, belonging to a certain group or social class can be expressed with brands in the sense of a status symbol. These factors are primarily shaped by the components values and visions on the part of the brand identity (Burmman et al. 2015).

Brand management therefore should identify consumer needs and address them by means of the benefits promised by the brand. Furthermore, it has to ensure that these promises are fulfilled by brand behaviour leading to a consumer experience which is in accordance with the promises and expectations (Burmman et al. 2015). Brand behaviour comprises both the brand's products and services as well as its communicative (and spatial) appearance – i.e. all aspects that shape the interaction between customer and brand at the various touch points.

The point of sale (PoS), i.e. the shop in the offline world, plays an outstanding role as a brand touch-point for consumers. While the importance of the point of sale for purchasing decisions is evident from a business point of view and has been researched in many different ways, the role of the PoS for

identity-based brand management is one of the aspects that has not yet been discussed and researched thoroughly. In the following, a brief look at the basic aspects of the point of sale will be shed before its role in the context of identity-based brand management will be discussed in detail.

2.2 Role of the point of sale

Within the customer journey, the point of sale is the touch point where the purchase actually takes place (see figure 1). For retailers, however, the relevance of the point of sale is not only determined by the fact that the purchase is made here; studies show that 70% of consumers just make their buying decision here. In addition, consumers tend to make impulse purchases – about 30% of purchase decisions are spontaneous and unplanned (GfK Gruppe 2009; Schröder 2015; Schröder et al. 2007; Gröppel-Klein/Bartmann 2007).



Figure 1: Customer journey

Source: own illustration

Due to this, the store layout is of crucial importance for the success of stores. Shop layout means the division and arrangement of the salesroom as well as its structuring into individual functional zones (merchandise area, customer area). With the store layout retailers decide which areas are reserved for the presentation of goods and what are the walking paths and, if applicable, the areas reserved for customers (e.g. special recreation zones in book stores etc.).

Quantitative space allocation deals with the size allocation of sales areas to product groups, while the arrangement of the individual product groups on the sales area is referred to as qualitative space allocation. The promotion of impulse buying plays a central role in this arrangement. These aspects are particularly important for the food, consumer electronics, cosmetics and fashion retail sectors, i.e. for products that belong to the categories of convenience or shopping goods.

However, with the exception of food, these are also those sectors in which the point of sale is partly shifting due to the development and expansion of e-commerce. Factors such as high transparency, time savings and convenience are increasingly influencing consumers' purchasing behaviour and ultimately lead to digital purchases or at least digital channels playing an important role in offline purchases in the sense of the 'omnichannel' concept (Westermann et al. 2018; Mehn/Wirtz 2018).

For stationary retailers, this means that targeted instruments must be used to keep the brick and mortar stores attractive. This includes the possibility of experiencing products live, personal service and advice from sales staff and social interaction with other consumers – but also the spatial experience of the product range and the brand in the sense of experiential marketing.

These experiences can be seen as a central factor for the formation of the brand image in the tradition of Experiential Marketing (Schmitt 1999). In contrast to rational decision theory with a focus on purely functional aspects, this approach assumes that decision makers always behave at the same time rationally *and* emotionally and strive for pleasant experiences. Schmitt distinguishes between five different types or levels of experiences, so-called strategic ‘experience modules’ (Schmitt 1999: 56ff; see figure 2).

First, sensual experiences (‘sense’) focus on sensory perception and appeal to the human senses. Second, the module ‘feel’ generates emotions and feelings. Third, creative cognitive experiences are covered by the intellectual level and encourage consumers, for example, to ‘think’ about the brand. Alternatively, brand knowledge can also be conveyed in this way. Fourth, the level ‘relate’ expresses the social and cultural context of a brand – a space is created for encounters with the brand community. The last level (‘act’) includes behaviour and lifestyle. Consumers can, for example, experience products live through product tests or get in contact with other people (Schmitt 2010).

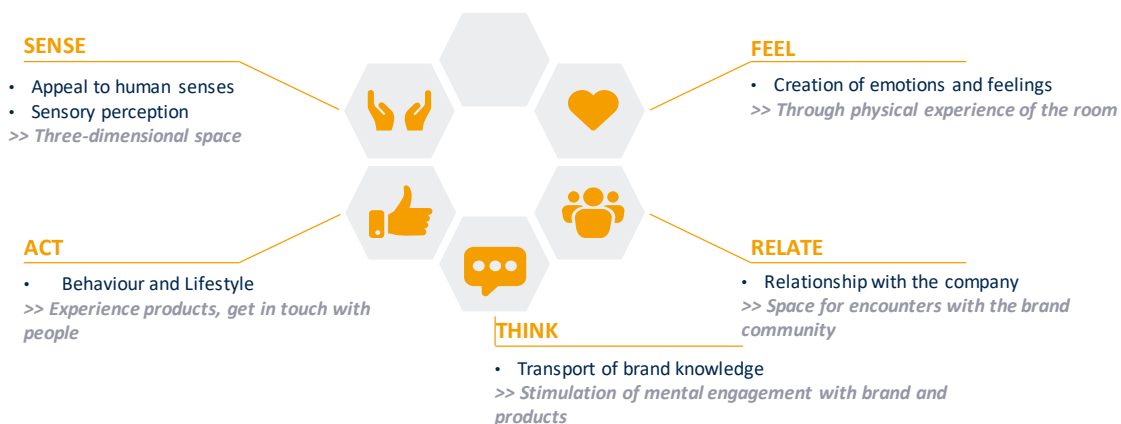


Figure 2: Experience modules

Source: adapted from Ghorbani/Westermann 2023 based on Schmitt 1999

These experiences are created by so-called experience providers such as communication and as well by the (interior) design of the point of sale (Schmitt 1999).

The aim is to create holistic experiences or to integrate personal experiences into a holistic framework, so that the places are not only decisive for the duration of the respective experience, but also serve as an anchor of memory for individually significant experiences. In this way, brand associations are positively influenced and – in the most positive case – an unforgettable and emotional connection between brand and customer is created (Raffelt/Meyer 2012).

Emotions associated with a brand, brand knowledge and values associated with the brand are also a crucial aspect in the context of identity-based brand management. Therefore, marketing, communication and sales measures should be implemented in such a way that positive and specific emotions are evoked in the consumer which are in line with the defined brand identity (Esch et al. 2012) – and the interior design of the point of sale plays a central role here.

In addition to the behaviour, the service and advice of the sales staff, the ambience in the stores is decisive for the consumer perception (Müller 2012; Steiner 2011). This can be influenced by the targeted, strategic use of a number of elements including music, scents and interior design. If brands are understood as constructs with a personality comparable to those of actual people, then shops in fact become ‘living rooms’ of these brand personalities and the way of living and furnishing is a central part to express one's personality.

Whereas authors have dealt with the influence of scents and music in the context of brand management, the role of interior design has been neglected by far. Thus, in the following, this paper focuses on interior design, discussing the different aspects in which it can be used from a theoretical point of view and what it can achieve. In this context, the authors refer to the broader term Architectural Branding after Raffelt (2012) and Raffelt/Meyer (2012). It comprises also brand related exterior design in addition to interior design.

2.3 Strengthening brand awareness, identity and experience through Architectural Branding

Architectural Branding is part of the broader concept of Corporate Architecture which is defined as ‘a means of communicating content and information about the respective company’ (Messedat 2005). Architecture is thus understood as a form of communication and offers companies the opportunity to present their values to internal and external target groups in a sustainable and continuous manner in the form of buildings. (Raffelt/Meyer 2012) In the corporate context, architecture includes all building concepts (e.g. production, administration or sales buildings) and (interior) space concepts (e.g. shops) but also temporary presentation concepts such as exhibition stands.

An essential goal of (Interior) Architectural Branding, as that part of Corporate Architecture which focuses on architectural elements of brand touch points, is to create these touch points in a way in which they strengthen brand perception positively and sustainably through a brand-compliant (interior) architectural appearance (Raffelt/Meyer 2012). Both brand awareness and identity can thus be supported by architectural elements. While distinctiveness and consistency contribute to the recognition of the brand and thus to awareness, the values and the personality can also be experienced visibly and sensually with the specific design. Thereby the wall, ceiling and floor design as well as the interior design are of central importance. In this way, a unique encounter with the brand and a high degree of authenticity can be achieved (Raffelt/Meyer 2012).

2.4 Dimensions of store design and their connection to brand identity

But how can brand values and brand personality be expressed by means of interior design elements from a theoretical point of view? To answer this question, it is necessary to consider the different relevant dimensions. The interior design comprises colours and forms used for the walls, the ceiling and the floor as well as decorative elements. All these elements are closely connected to the presentation of goods and will be considered more in detail.

In a broader sense also visual, acoustic and olfactory shop environment factors belong to it. Exterior design also comprises architectural (design) elements equivalent to the interior design, which can only be perceived from the outside (e.g. shop front design). If the store is not rented but newly constructed, the whole building obviously belongs to the exterior design (like e.g. in the case of Café del Sol where all cafes are newly build helping to express the brand identity).

By combining various coordinated design elements, a certain store atmosphere can be created which results from the sum of sensory impressions of the consumer (Theis 1999). The influence of this perception on the brand image depends on the extent to which the retailer manages to communicate the defined brand values through corresponding elements. In order to assess the potential impact of store design and atmosphere on consumer's brand perception, it is crucial to consider both the relevance of individual design aspects and consumer's overall impression. For this reason, this study focuses on (interior) architecture and visually perceptible shop environment factors. Interior and partially exterior design are therefore of particular interest. The individual design parameters of interior and exterior design can now be further differentiated according to five similar categories (see figure 3).

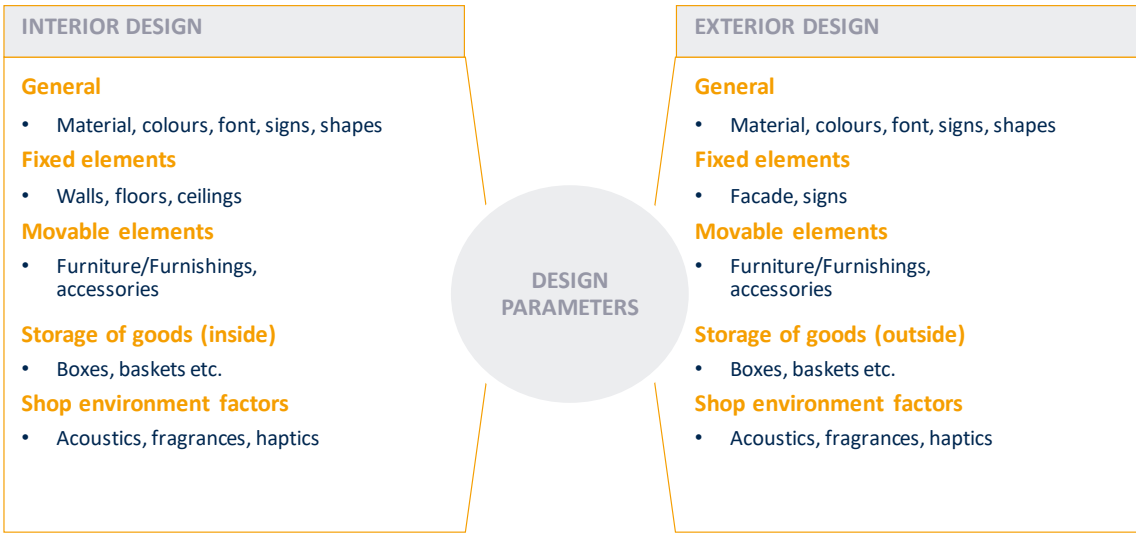


Figure 3: Design parameters
Source: own illustration based on different authors

First of all, the materials, the colours, signs, fonts and shapes used play a crucial role and should be taken into consideration. With regard to the five basic types of Aaker (1997) mentioned at the beginning, brands which have chosen an exciting character would – so the assumption – choose arousing colours like red, while brands with a cultivated character would tend to use more decent colours. But also, forms can be used to express character traits: Cultivated brands are presumably round in shape, while 'robust' trademarks would be more likely to use square forms with edges.

In addition, the fixed elements of interior design should also be considered, i.e. walls, floors and ceilings. Brands that stand for diversity can use different, alternating design features, whereas brands that stand for clarity and stringency are more likely to tend towards uniform, recurring design variants. Also, movable elements, i.e. furnishings such as furniture, accessories and decorations, must also be considered from similar points of view. The same applies to the form and method of storing goods. Furthermore, additional shop environment factors like scents, sounds and the haptic of the materials used in the store also contribute to the brand perception and can be realised in accordance with the brand identity, e.g. rough brands would use rough materials, masculine scents and arousing music, whereas cultivated brands should stick to smooth haptic impressions, neutral to feminine scents and calming sounds.

In order to investigate the extent to which retailers succeed in conveying their brand identity within the framework of shop design and actually turn the shop into the living room of the brand, all design categories must be considered. It is important to consider which individual design parameters are used in interior and exterior design and to what extent they are related to the brand identity – as has been

explained by the examples mentioned. Although it is crucial to address consumers via more than one sensory channel within the framework of holistic brand communication, visual perception dominates (Müller 2012). Accordingly, the following explanations concentrate on the interior design of retail stores and thus the visual dimension.

In order to systematically consider how the personality and values of the brand are expressed through the interior design of the stores, all design parameters must be analysed with regard to their connection to personality and values. This connection can be established by an analysis grid in matrix form. The grid consists of the components of the brand identity as rows and the categories of the design parameters as columns (Westermann/Zimmermann 2019a). This results in the following system:

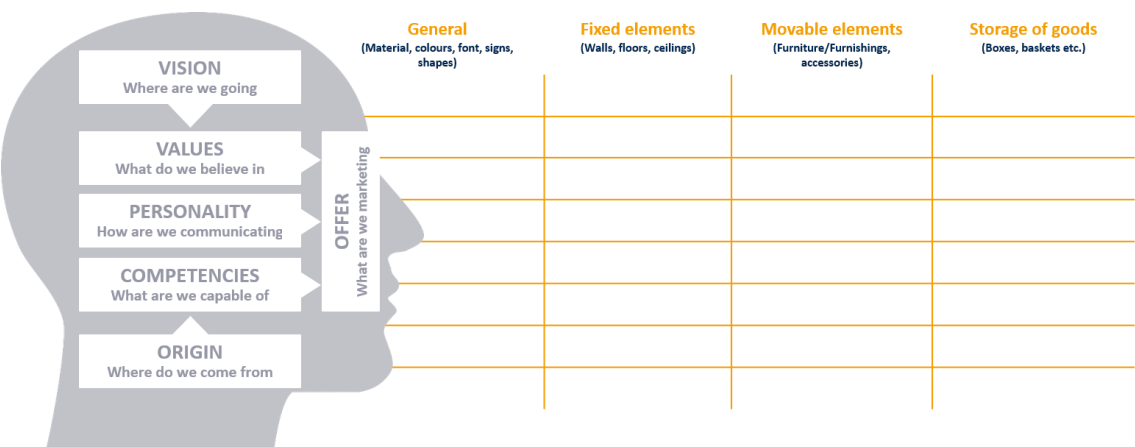


Figure 4: *Analysis grid*
Source: *own illustration*

2.5 Specialities in the creative industry

Especially for companies in the creative industry, there are many opportunities to express their brand identity at the PoS and anchor it in the minds of their stakeholders. For example, cooperation with artists and the integration of art at the point of sale is a win-win situation for both sides: for artists, exhibitions in stores and brand spaces or commissioned works for concrete spatial situations provide an opportunity to promote their own work. For companies, investing in art that matches the brand identity is much more than just a decorative addition to the working environment or the store: it ensures that the identity becomes vividly tangible and is also a sign of cultural commitment that can lead to a creative dialogue both among the company's own employees and externally among customers (Conzen/Salié 2012). Hybrid architectural concepts offer new perspectives for the encounter between customer and brand.

Such shared spaces correspond to the principle of Design Thinking, whose basic idea is to bring together as many different experiences, opinions and perspectives as possible. In contrast to other innovation methods, Design Thinking is described as an approach based on the three basic elements team, space and process. The principle of sharing creative spaces and processes opens new perspectives for the development of brands and the expression of their personality. In addition to static spaces, dynamic spatial situations such as pop-up stores can be used for exciting encounters between brands and art. A selected work can be integrated here or even the store itself becomes a temporary work of art (Conzen/Salié 2012). In this way, the brand identity becomes alive and tangible.

3 Materials and methods

3.1 Research gap and research aim

As discussed in the literature review, Corporate Architecture plays a central role in communicating brand identity, as it creates a brand environment and an interactive space for consumers and employees. In addition to illustrating corporate design or strengthening brand recognition, Architectural Branding aims to enhance brand-congruent consumer experiences. However, existing research concerning the point of sale focused on different sales aspects. Existing studies deal for example with the importance of walking routes, the placement of goods and the use of displays (Schröder et al. 2007; Gröppel-Klein/Bartmann 2007; Berekoven 1990; Theis 1999). In addition, there are a number of studies dealing with multisensory factors such as scent or sound marketing (e.g. Müller 2012). A pure consideration of the interior design dimensions as a form of expressing brand identity, has hardly been the subject of previous research or considerations of its significance for identity-based brand management. Although some theoretical considerations with regard to Architectural Branding and Corporate Architecture have been made in recent years, most studies have focused on exterior architecture (Messedat 2005; Raffelt 2012; Raffelt et al. 2013; Vonseelen 2012; Brauer 2011; Schröder 2015). By contrast, interior architecture and related design choices remained a research gap to a large extent (Leydecker 2014; Dagli 2016; Shon/Song 2021).

The authors aim is to provide an initial approach to fill this gap, which leads to the following key research questions:

1. How do brands and retailers use interior design to transport the brand identity and which concepts can be observed?
2. What influence do these concepts have on the perception of consumers?

In order to answer these research questions, an explorative study design consisting of two parts was developed, which comprises a qualitative and a quantitative methodology with the aim of evaluating the interior design of the store and its impact on consumers. Due to the explorative character of the

study, the authors decided to focus on one suitable case. Thus, in a first step a suitable brand had to be identified. In the second step, the stores of this brand were analysed in a qualitative way on the basis of the theoretical findings by categorising and interpreting store design parameters with the help of the theory-based analysis grid developed in the theoretical chapter. This study part delivered answers for the first research question. Afterwards the authors elaborated a survey design for testing the impact of the store design on consumer perception of the brand taking the experiences into regard which had been made in an online survey among students (Westermann/Zimmermann 2019b). Based on the results and methodological experiences made during this study the methodology and the questionnaire have been developed further. Finally, an online survey representative for the German population between 18 and 64 was carried out to answer the second research question and to test the main hypothesis that values can be transported via means of interior design. The results of this survey will be the further focus of this paper.

3.2 Brand selection and qualitative case study

After a comprehensive pre-search for suitable brands, the brand Rituals was selected as the research object for the case study and the subsequent consumer survey. At the beginning of the case study process, several shops in different European countries were visited and the following two aspects were identified as prerequisites for further research. First, the stores are designed identically, independent of the country and the city where the store is located. Second, most stores have approximately the same size, thus they are structured in the same way. It may be concluded that Rituals employs a strategic, cross-national approach of communicating a consistent brand identity at the point of sale.

Rituals has been on the market for 23 years and may therefore serve as an example and blueprint of how a brand can be successfully established on a B2C mass market in the 21st century. Today the brand has over 600 brand stores and numerous other sales outlets in 27 countries (Rituals 2018).

Rituals defines itself as a luxury lifestyle brand and as the first brand to combine luxurious home and body care products with the aim of transforming everyday routines into meaningful rituals. The philosophy of the brand is that happiness lies in the detail and the small and simple things that need to be enjoyed with care and attention. Ritual's understands its brand and products as inspired by Far Eastern traditions, being of high quality and meant to enrich body and soul. At the same time, they should remain affordable due to a fair price-performance ratio. This already leads to the identification of brand values and characteristics, which is important as they can serve as quantifiable items for the consumer survey. Therefore, the company's website and printed material were analysed, and additionally an interview was conducted with the brand's marketing manager. This led to the identification of the following five key brand attributes: 'luxurious', 'unique', 'far eastern inspired', 'affordable' and 'natural'.

3.3 Qualitative store check

The Rituals stores usually have an area of only 60-100 square meters. The overall design of the stores is evocative of luxurious wellness spas. The basic aim here is obviously to provide consumers with a pleasant and relaxed shopping experience. Various design elements can be identified in the individual categories of the analysis grid mentioned in the theoretical part to express the defined brand values.

By using relatively dark colours (e.g. black lacquer) and dark materials, above all (solid) wood, as well as dimmed light, a pampering, relaxing and luxurious atmosphere is created throughout the entire store immediately upon entering. In addition, the various wooden elements are meant to create a feeling of sustainability, authenticity and high quality. The same counts for the floor and wall design: The dark parquet floor has a high-quality effect. Parquet elements can also be found partly on the walls in addition to stucco realised with a special wiping technique. This mixture underlines the brand characteristics of variety and, again, of quality and luxury.

The relaxing and pampering spa ambience is further enhanced by various accessories. These include, for example, beach mats, curtains, relaxing pictures and photo motifs, various grasses but also daylight mirrors.

Far Eastern flair is created by a Buddha statue and white jasmine. The dark (mainly black) colour also shows Far Eastern inspiration.

Products are mainly presented in black lacquered shelves with glass fronts and in bast baskets. Also, clothes racks are used and occasionally the products are also presented simply on tables. The shelves look high-quality and luxurious. Baskets and coat racks convey sustainability, authenticity and relaxation as well as Far Eastern inspiration.

Due to the brand values of Rituals, it is noticeable that a limited choice of colours as well as clear forms and structures are used in the shop design. The relatively clear focus of the brand values and their matching obviously facilitates a stringent design of the PoS.

3.4 Specification of the second research question and the according hypotheses

Following this preliminary qualitative investigation, the second central research question was specified: How does the (interior) architectural design of Rituals influence consumer's perception of the single brand values and attributes? Our major hypothesis (H1) is, that being confronted with the interior design at the point of sale leads consumers to attach the desired values and attributes to the brand. In a representative survey presumably not all respondents know the brand before being confronted with it during taking part in the survey. Nevertheless, the effect of transporting the intended brand values and attributes should work for both groups (H2a), but the attribution effect among the

group of people who have known the brand before should be stronger than in the other group (H2b) as they have had brand knowledge before which is just actualised during the survey. In accordance with this, there should also be an influence of the point in time when consumers knowing the brand have visited a Rituals store for the last time: If the last visit is not long ago the brand values and attributes should be more present even without being exposed to the store design, whereas a longer period between the last visit and taking part in the survey leads to a weakening of attached brand attributes (H3a). This effect should be partially levelled after being exposed to the store design (H3b).

Finally, research shows that the different senses are of different importance for people. It can be assumed that people who are of the opinion that visual impressions are of importance to them are more likely to be influenced by a primarily visual store concept than people who stress the importance of other senses (H4).

3.5 Quantitative survey design and sample

A quantitative, self-completion online survey was designed to measure the perception and strength of brand values and. In order to assess the relevance of interior design at the point of sale and to measure possible changes in brand perception, a simulation of a store visit was conceptualized taking the central research question and the different hypotheses into regard. Thus, photo-based questions were used in the questionnaire to give the respondents an impression as realistic as possible of the interior store design, which is a common practice in interior design studies (Petermans et al. 2014).

For the realisation of this study an online panel (omnibus) from a market research institute was used, in which 1,000 consumers from Germany were surveyed. The selection of respondents was representative for the online population between 18 and 65. The sample was representative with respect to gender and age¹ (see figure 5).

Initially, the authors asked people if they know the brand Rituals. If they agreed they were asked when they visited a store for the last time. Furthermore, the respondents were asked if they usually recognise visual aspects, music and scents when entering a store.

The study participants knowing the brand were asked to evaluate how strongly they attach brand attributes on a 5-point Likert scale both before and after viewing photos from a Rituals store. Respondents not knowing the brand were directly exposed to the store photos and then asked how they perceive the brand.

It is important to note that the items of this scale encompassed both the intended, but also two not intended brand attributes. On the one hand, the authors thus examined whether the five intended

¹ A deeper analysis of further details and the impact of sociodemographic aspects would have been interesting, however due to financial restrictions and a limited budget this was not possible. Instead, the focus of the study was comparing participants based on their previous brand knowledge.

brand attributes were conveyed by design of the stores. On the other hand, they tested whether the three non-intended, inappropriate brand values may also be perceived to cross check the reliability of the survey results.

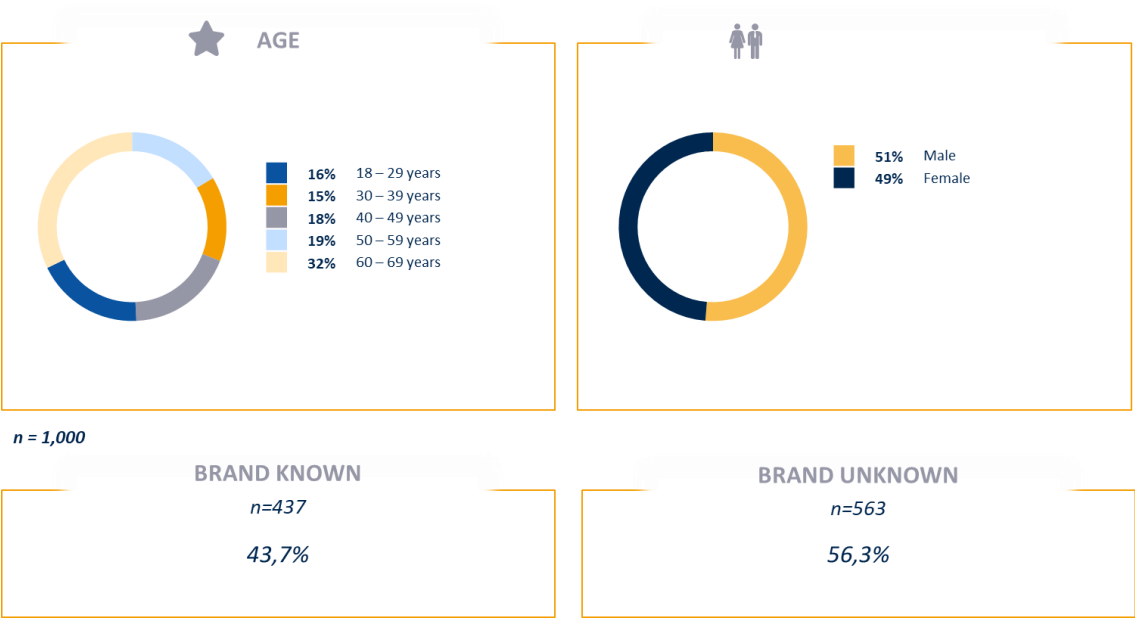


Figure 5: Sample overview
Source: own illustration

4 Results

4.1 Relevance of interior design elements and importance of visual aspects

Before elaborating on the ratings of brand values and attributes, the authors asked the participants to rate their conscious perception of interior design parameters with 4- or 3-point rating scales. This served the purpose of determining to what extent the respondents perceive interior design and atmosphere when entering a shop. In general, the majority of those surveyed stated to consciously perceive the fixed and movable elements of interior design as well as atmospheric elements. Most survey respondents stated that they perceive the furnishings (79%) and decoration (76%) directly. Walls (63%) as well as floors and ceilings (59%) seem less conspicuous in comparison. In addition, this enquiry showed that the ambience is perceived very strongly through the aspects of scents (84%), music (80%) and light (78%). A notion of caution is due here, since this survey only indirectly measures the effect of the interior design based on the participants' self-disclosure. Therefore, unconscious influences and their impact remain uncertain.

Based on these observations on the relevance of interior design elements, the entire sample was divided into two types to find out if there are customers that are more influenced by visual aspects. To this end, it was determined which customers expressed a high level of approval for all four visual aspects. The comparison of these visual types with non-visual types did not yield significant results in terms of brand evaluation. There were slight differences with regard to the fact that visual types overall perceived brand values more strongly. However, the impression of the photos and the resulting differences in image ratings did not show clear differences between the types. Thus, it can be concluded that hypothesis H4 must be rejected, according to which visual types are more influenced by the visual store experience in their brand evaluation.

4.2 Brand awareness, last store visit and brand perception

43.7% of the participants stated to know the brand. Also, 29% of the total sample visited the store in the last six months prior to the survey. In contrast, 56.3% said they did not know the brand. In the following, the statistical analysis distinguishes between the brand perceptions of these two groups to determine possible differences in the perception of interior design and the assessments of brand values.

The participants who stated that they knew the brand ($n = 437$) were first asked to assess how strongly they agreed with the eight attributes with regard to the brand before viewing images of the interior design of the Rituals stores. It is important to note that about half of the interviewees agreed with the five intended brand value items. For instance, 47% of this partial sample (rather) agreed that Rituals is 'luxurious' and 49% evaluated the brand as 'far eastern inspired'. Moreover, the brand value item 'natural' was most strongly perceived, as it received the approval of 58% of the sample. However, 53% of those surveyed perceived the brand to be 'exciting' and 'summerlike', which are considered as not intended brand values. The lowest agreement received the not intended attribute 'daring' with 29%.

The subjects were then shown two photos from the interior of a Rituals store and were asked to evaluate the atmosphere conveyed using the various brand attributes. The first photo elicited strong agreement with the atmospheric values of 'luxurious', 'unique' and 'far eastern inspired'. The perception of 'luxurious' as a dominant attribute was highlighted even more clearly, with 68% of the respondents agreeing. It is crucial to highlight that each of these photo attributions exceeded the previous scores for the overall brand attributes with an increase of at least 10%.

The same procedure was carried out for the second photo, which depicted the viewing angle from the entrance to the store. Similar to the first assessment, this photo was rated 'luxurious', 'unique' and 'far eastern inspired'. Each of these attributes also received at least 10% more approval than the overall brand rating before. However, the attribute 'unique' (67%) was rated most highly in this case. In addition, viewing this photo strengthened the attribute 'daring', what may have led to an unintentional brand impression.

Finally, participants were asked to re-evaluate the brand after viewing the photos. This question did not directly follow the exposure to the two photos, but followed later in the omnibus survey, i.e. questions concerning other topics were prior to the re-evaluation. In this re-evaluation, the three attributes ‘unique’ (66%), ‘luxurious’ (62%) and ‘far eastern inspired’ (62%) were considered as the most dominant brand values. Their individual increase of at least 10% shows that the impression of the photos has thus been transferred to overall brand evaluation. Also, the not intended feature ‘daring’ received an increase of over 10% with a final score of 39%. Interestingly, the previously most highly rated aspect ‘natural’ was no longer included in the top attributes. Another interesting aspect is that all items regardless of the brand’s intention have increased after viewing the photos (see figure 6).

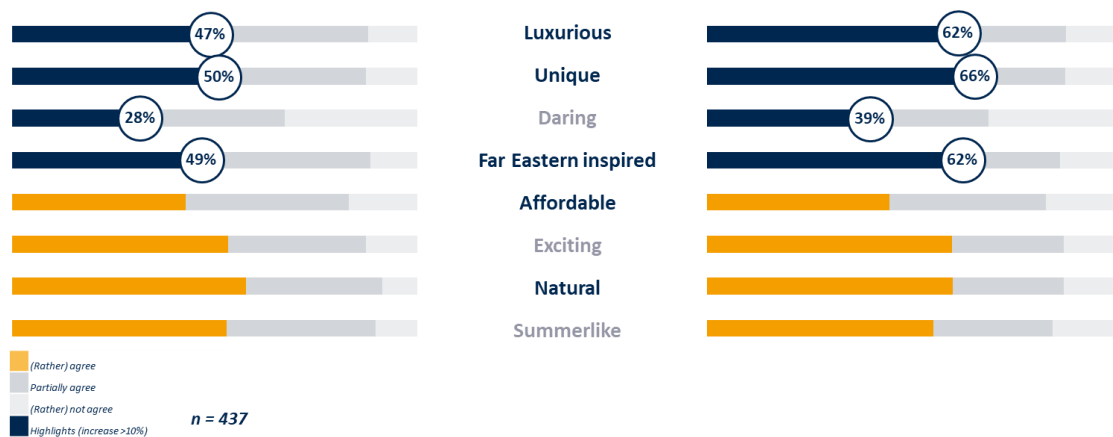


Figure 6: Image attribute comparison (before vs. after)
Source: own illustration

In contrast, the partial sample of participants without prior brand knowledge (n = 563) was – as already outlined – not asked to evaluate the brand at the beginning as this would have made no sense. They started with viewing the photos immediately. With reference to their evaluation of brand attributes as conveyed by the first photo, it becomes evident that they generally rated the attributes lower than respondents who knew the brand. This is illustrated by a comparison of the three top attributes. Although the strongest attributes here are identical to the top attributes of the first sample, the individual scores for the items ‘unique’ (44%) and ‘far eastern inspired’ (35%) are significantly lower. The impression ‘luxurious’ is the only exception with a value of 61% of agreement. Likewise, the rating of the second photo revealed the same three top attributes, albeit again with lower scores than in the first sub-sample. Again, the rating of the attribute ‘luxurious’ stands out with a score of 58%.

These findings are also reflected in the final brand evaluation after viewing the photos, with the attribute ‘luxurious’ being the only item that more than half of the test participants agreed to (55%). This is followed by ‘unique’ (41%) and ‘far eastern inspired’ (39%). The intended brand attribute ‘affordable’, on the other hand, only reached a value of 14% and is rated significantly worse than all unintended

brand attributes. In summary, it has been again shown that the attributes of the photos were transferred to the overall brand rating. The findings provide evidence for the main hypothesis H1 as they confirm the connection between interior design impression and brand value attributions. Nonetheless, the lower overall scores indicate that the effect is lower with regard to participants without previous brand knowledge as had been claimed in hypotheses H2a and H2b. (see figure 7)

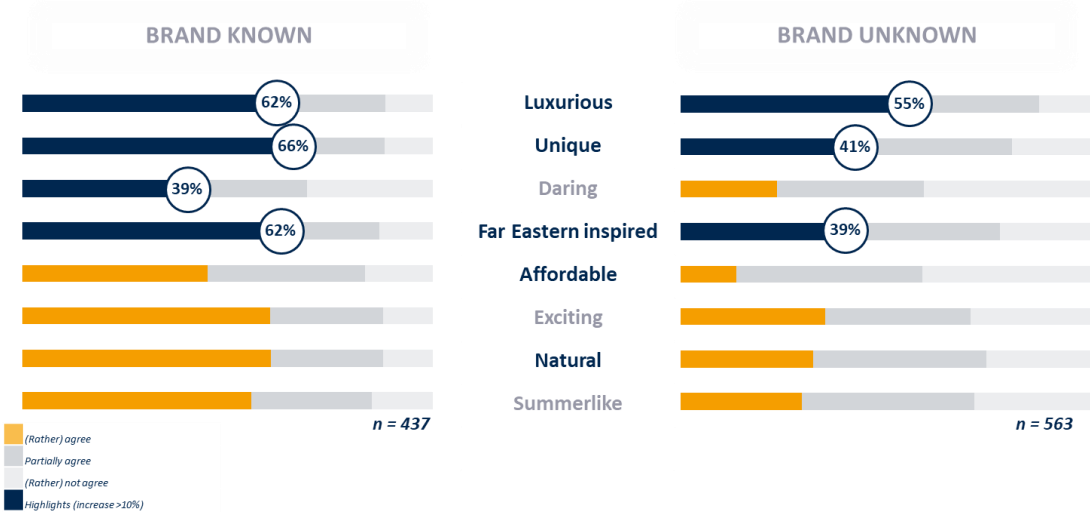


Figure 7: Consumer sample comparison (brand known vs. brand unknown)
Source: own illustration

What is more, the partial sample of consumers who knew the brand was further subdivided into visitors who have visited a store recently, that is within the last sixth months (n = 284), and those, whose visit was more than six months ago (n = 153). A comparison of these conditions showed that there are differences in the brand evaluation both before and after viewing the photos. First, the initial brand rating of the recent visitors was higher in all aspects than that of those who had visited Rituals longer ago. Moreover, the chi-square independence test confirmed the statistically significant correlation for all image items between the two groups. These results are significant at the $p = 0.05$ level. With respect to the intended brand values, the biggest differences were found for the attributes ‘far eastern inspired’ (56% vs. 37%), ‘affordable’ (50% vs. 29%) and ‘natural’ (64% vs. 46%). However, these correlations also apply to the not intended brand values. Therefore, these correlations should be interpreted with caution. Nevertheless, they show that the overall brand impression appears to be stronger.

It can be inferred therefrom that consumers who visited the store recently are more likely to perceive the brand as desired. This can be attributed to the brand experience in the store, which is strongly influenced by the implementation of the brand identity at the PoS. If the visit has taken place some time ago, consumers are more reluctant to agree with the brand values, what may be traced back to a weaker recollection of the brand experience due to the time passed in the meantime.

Similarly, the brand evaluation after viewing the photos was investigated. This also showed that the agreement of those who were recently in a store is higher in all points, but the differences are not as pronounced as before. The statistical independence test also only partially shows correlations between the visitor groups, namely for the attributes 'luxurious', 'unique', 'affordable', 'exciting', 'natural' and 'summerlike'. Basically, the approval rates for both groups have increased after viewing the photos, which indicates that the existing brand knowledge is actualised by viewing the photos. In both cases, the attribution of the desired brand values is increased. For those who have not been in the store for a long time, there is a greater update effect and accordingly the gap to visitors with more recent experiences is reduced.

Taken together, the two hypotheses H3a and H3b were confirmed, since the point at time at which consumers last visited the store significantly influenced their brand knowledge and the potential strength of brand value impressions. Possible differences, however, were partially relativized by the impression of photos. The results of this quasi-experimental comparison based on visual stimuli can also be seen as evidence for the assumption that the visual impression is a central criterion for the brand-congruent interior design.

5 Conclusion

The theoretical discussion and the additional qualitative store check have shown that there is a large range of possibilities to express the brand identity with its values and its personality at the PoS. The brand Rituals chosen for this case study shows that many of these possibilities are used in brand management practice in this case.

The quantitative study set out to investigate whether and how interior design influences the perception of brand values and attributes taking the same case as example. Using photo-based questions that simulated the visit to a Rituals store, the online survey measured the brand perception of consumers. This study has shown that interior design and atmosphere have a high impact on the perception of brand values and attributes. On the one hand, the quasi-experimental study showed that interior design can actualise and strengthen intended brand values in relation to the perception of consumers who know the brand. On the other hand, it pointed out that among those who do not know the brand, the newly created images tend to follow the intended direction in terms of brand values. The intended value 'affordable' is the only brand value which seems not to be transported by the interior design of Rituals. This is not surprising, as this result is in accordance with the store check which revealed that the brand places its focus at the PoS on the also intended value 'luxurious'. Both attributes stand in contrast to each other as far as visual appearance is concerned and form a general challenge for the management of this brand.

With reference to the model of identity-based branding as the starting point for Corporate Architecture and Architectural Branding, this research supported the idea that not only exterior architecture, but also interior architecture contributes significantly to brand recognition and evaluation. The findings reported here shed new light on the communication and action potential of interior design, as the store seemed to be considered as the living room of the brand expressing its personality.

The results underline the importance of brand identity implementation at the point of sale by showing that consumers consciously perceive interior design and atmospheric elements. Moreover, the findings suggest that these impressions are related to the formation of brand images, as they may enhance or even change the strength of brand value attributions. The insights gained from this study may be of assistance to retail companies that aim to enhance brand-congruent consumer experiences. Finally, this research suggests that brand identity perception by consumers with and without previous brand knowledge can be achieved through interior design if the design elements underline the intended brand values.

5.1 Practical Implications

The results suggest that store design should consequently be in line with the brand identity, its values and personality. To achieve this, the implementation of the brand identity should be a part of the conceptual work with regard to store concepts in addition to classic aspects like designing the layout and consumer walking paths. It is crucial that this conceptual work is done from a holistic perspective taking all dimensions into account, beginning with the materials and forms used, through the design and integration of floor, ceiling and walls into the concept, to decorative elements and signs. The analytical grid developed at the end of the theoretical part may help to consider all relevant dimensions. Finally, market research pre-tests of newly conceptualised stores should also consider the influence of the store design on the brand perception and should not be limited to sales aspects.

The design of the PoS (and of company buildings in general) offers great opportunities to introduce their brand identity to consumers and other stakeholders: Welcoming them in the brand's living room may leave a lasting first impression establishing a brand image which is in line with the brand identity. Moreover, companies from the creative sector can express their creativity right here choosing new and unusual ways of conceptualising their 'living rooms'.

5.2 Limitations and future research

The methodological approach used in this study simulates a store visit by showing two pictures of Ritual stores from different angles. Naturally, the two pictures do not give a holistic impression of the store, as not every detail and every possible perspective can be seen, and furthermore the pictures are limited to two dimensions thus not really giving a completely realistic impression of a 3D room.

Moreover, music and sounds as well as scents and the haptic of the materials used cannot be transmitted. Respondents have to evaluate on the basis of an incomplete slice of reality. The authors accepted these limitations in order to obtain representative data.

Although a general influence of interior design at the PoS on the brand perception can be claimed, results may not be generalised in detail with regard to the statistical data, since the study presented in this article is limited to the case of one company. Furthermore, as far as the qualitative part is concerned, the analytical grid is primarily a blueprint for a criteria-based analysis. Future research of the ways the brand identity is implemented at the PoS could use more elaborated, psychology-based analytical tools.

Future research should also be devoted to test the effects identified here with a much broader basis of brands from different industry sectors and their respective stores, both to check if the effect is the same in other sectors like food, electronics, clothing etc. or if it is dependent on the type of goods and the resulting predisposition of customers. Such a broader basis of brands could help to generate detailed data which can be generalised thus delivering reliable data to quantify the effect.

Furthermore, in addition to approaches using simulations, research should also be devoted to field studies in which subjects are interviewed directly after visiting stores. This could either be realised with a qualitative approach with in-depths interviews or as a quantitative approach. For such a quantitative approach the cooperation with a brand would be ideal, as it would allow either interviews directly in or in front of the store or even the use of customer data collected in loyalty programs.

Finally, the effect identified here could be analysed in more detail as far as different types of customers are concerned. The study did for example not focus on the influence of age or gender with regard to the research question. And though the hypothesis H4 could not be approved, it still remains open if the effect is different in dependence from the predisposition and type of the consumer, as the limited number of questions which can be asked in an omnibus survey did not allow to integrate questions which may have led to a sophisticated differentiation with regard to these aspects.

References

- Aaker, D. A. (1996): Building strong brands. New York: Free Press.
- Aaker, J. L. (1997): Dimensions of Brand Personality. In: Journal of Marketing Research, vol. 34 (1997), No. 3, pp. 347–356.
- Bauer, H. H.; Heinrich, D.; Samak, M. (eds.) (2012): Erlebniskommunikation. Erfolgsfaktoren für die Marketingpraxis. Berlin: Springer.
- Berekoven, L. (1990): Erfolgreiches Einzelhandelsmarketing. Grundlagen und Entscheidungshilfen. München: Beck.

- Böckenholt, I.; Mehn, A.; Westermann, A. (eds.) (2018): Konzepte und Strategien für Omnichannel-Exzellenz. Innovatives Retail-Marketing mit mehrdimensionalen Vertriebs- und Kommunikationskanälen. Wiesbaden: Springer Gabler.
- Brauer, G. (2011): Architektur als Markenkommunikation. Basel: Birkhäuser Verlag GmbH.
- Burmann, C.; Halaszovich, T.; Schade, M.; Hemmann, F. (2015): Identitätsbasierte Markenführung. Grundlagen – Strategie – Umsetzung – Controlling. 2nd ed., Wiesbaden: Springer Fachmedien.
- Burmann, C.; Halaszovich, T.; Schade, M.; Piehler, R. (2018): Identitätsbasierte Markenführung. Grundlagen – Strategie – Umsetzung – Controlling. 3rd ed., Wiesbaden: Springer Fachmedien.
- Burmann, C.; Riley, N.-M.; Halaszovich, T. F.; Schade, M. (2017): Identity-based brand management. Fundamentals – strategy – implementation – controlling. Wiesbaden: Springer Gabler.
- Conzen, F. G.; Salié, O. (eds.); Kulturkreis der Deutschen Wirtschaft (ed.) (2012): Corporate collections. [feat. Ritter, Deutsche Bank, AXA, Faber Castell, Allianz, WS, B.Braun Melsungen, Gen RE, Mont-blanc ...]. Köln: Deutsche Standards EDITIONEN.
- Dagli, Z. (2016): Corporate identity in interior design. In: Global Journal on Humanities and Social Sciences, vol. 4 (2016), No. 3, pp. 318–327.
- Esch, F.-R.; Gawlowski, D.; Rühl, V. (2012): Erlebnisorientierte Kommunikation sinnvoll gestalten und managen. In: Bauer, H. H.; Heinrich, D.; Samak, M. (eds.): Erlebniskommunikation. Erfolgsfaktoren für die Marketingpraxis. Berlin: Springer, pp. 13–30.
- Esch, F.-R. (2018): Strategie und Technik der Markenführung (9th ed.). München: Franz Vahlen.
- GfK Gruppe (ed.) (2009): STORE EFFEKT. Nürnberg: GfK SE.
- Ghorbani, M., & Westermann, A. (2023): Integrierte Markenführung im digitalen Zeitalter: Bedeutung – Konzepte – Instrumente – Evaluation. Wiesbaden: Springer Gabler.
- Gröppel-Klein, A.; Bartmann, B. (2007): Anti-Clockwise Or Clockwise? The Impact of Store Layout on the Process of Orientation in a Discount Store. In: European Advances in Consumer Research, vol. 8 (2007), pp. 415–416.
- Kapferer, J.-N. (1997): Strategic Brand Management: Creating and sustaining brand equity long term (2nd ed.). London: Kogan Page.
- Leydecker, S. (2014): Corporate Interiors. Grundlagen, Komponenten, Beispiele. Stuttgart: avedition.
- Meffert, H.; Burmann, C.; Koers, M. (2005): Identitätsorientierte Markenführung und praktische Umsetzung. 2nd ed., Wiesbaden: Springer Fachmedien.
- Mehn, A.; Wirtz, V. (2018): Stand der Forschung – Entwicklung von Omnichannel-Strategien als Antwort auf neues Konsumentenverhalten. In: Böckenholt, I.; Mehn, A.; Westermann, A. (eds.): Konzepte und Strategien für Omnichannel-Exzellenz. Innovatives Retail-Marketing mit mehrdimensionalen Vertriebs- und Kommunikationskanälen. Wiesbaden: Springer Gabler, pp. 4–35.
- Messedat, J. (2005): Corporate Architecture: Entwicklung, Konzepte, Strategien. Stuttgart : avedition.
- Müller, J. (2012): Multisensuale Gestaltung der Ladenatmosphäre zur Profilierung von Store Brands. Ein theoriegeleitetes, experimentelles Design zum Shopperverhalten. Wiesbaden: Gabler Verlag (Innovatives Markenmanagement, vol. 39).
- Petermans, A.; Kent, A.; van Cleempoel, K. (2014): Photo-elicitation: Using photographs to read retail interiors through consumers' eyes. In: Journal of Business Research, vol. 67 (2014), No. 11, pp. 2243–2249.

- Raffelt, U. (2012): Architectural Branding: Understanding and Measuring its relevance for Brand Communication. München: Fördergesellschaft Marketing Verlag.
- Raffelt, U.; Meyer, A. (2012): Wie die Marke in Architektur erlebbar wird. In: Bauer, H. H.; Heinrich, D.; Samak, M. (eds.): Erlebniskommunikation. Erfolgsfaktoren für die Marketingpraxis. Berlin: Springer, pp. 207–236.
- Raffelt, U.; Schmitt, B. H.; Meyer, A. (2013): Marketing function and form: How functionalist and experiential architectures affect corporate brand personality. In: International Journal of Research in Marketing, ABI/Inform Collection, vol. 30 (2013), No. 3, pp. 201–210.
- Rituals (2018): About Rituals (<https://www.rituals.com/de-de/about-rituals.html>). Accessed on 13.07.2019.
- Schmitt, B. (2010): Experience Marketing. Concepts, Frameworks and Consumer Insights. In: Foundations and Trends® in Marketing, vol. 5 (2010), No. 2, pp. 55–112.
- Schmitt, B. H. (1999): Experiential marketing. How to get customers to sense, feel, think, act, and relate to your company and brands. New York, NY: Free Press.
- Schröder, H.; Möller, N.; Zimmermann, G. (2007): Die Analyse des Such- und Entscheidungsverhaltens von Kunden im stationären Einzelhandel – ein Vergleich ausgewählter Beobachtungsmethoden. In: Ahlert, D.; Olbrich, R.; Schröder, H. (eds.): Shopper Research – Kundenverhalten im Handel. Jahrbuch Vertriebs- und Handelsmanagement 2007. Frankfurt am Main: Deutscher Fachverlag GmbH (Edition Lebensmittel Zeitung), pp. 156–172.
- Schröder, T. (2015): Corporate Architecture: Identität durch Architektur. Münster: Deutscher Architektur Verlag.
- Shon, J. H.; Song, J. (2021): Effect of Design Elements and Brand Identity on Brand Association, Brand Personality, and Brand Image: Focusing on Franchisee Coffee Shop. In: The Korean Journal of Franchise Management, vol. 12 (2021), No. 3, pp. 21–34.
- Steiner, P. (2011): Sensory Branding. Grundlagen multisensualer Markenführung. Wiesbaden: Springer Fachmedien.
- Theis, H.-J. (1999): Handels-Marketing. Analyse- und Planungskonzepte für den Einzelhandel. Frankfurt am Main: Deutscher Fachverlag.
- Vonseelen, T. (2012): Von Erdbeeren und Wolkenkratzern. Corporate Architecture – Begründung, Geschichte und Ausprägung einer architektonischen Imagestrategie. Oberhausen: Athena Verlag.
- Westermann, A.; Wirtz, V.; Zimmermann, R. (2018): Ein theoriebasierter Ansatz zur Erlangung von Exzellenz im Omnichannel-Vertrieb. In: Böckenholt, I.; Mehn, A.; Westermann, A. (eds.): Konzepte und Strategien für Omnichannel-Exzellenz. Innovatives Retail-Marketing mit mehrdimensionalen Vertriebs- und Kommunikationskanälen. Wiesbaden: Springer Gabler, pp. 37–70.
- Westermann, A.; Zimmermann, R. (2019a): Brand identity at the point of sale. In: Messedat, J. (ed.): Retail Design International. Components, Spaces, Buildings. 4th ed., Stuttgart: avedition, pp. 8–13.
- Westermann, A.; Zimmermann, R. (2019b): Innenarchitektur am Point of Sale: Wie Marken ihre Identität im Store zum Ausdruck bringen. In: Bielzer, L. (ed.): Corporate Architecture. Künzelsau: Swiridoff, pp. 57–67.

Authors

Arne **Westermann**, Ph.D., born in 1972 in Bochum/Germany, is Professor for Communications and Marketing at the International School of Management (ISM) in Dortmund. He is the Program Director for the Master Program Strategic Marketing Management and Head of the Brand & Retail Management Institute @ ISM. Additionally, he is in charge of Marketing in the ISM's educational division. Before joining the ISM he was Professor of Communication Management for the Quadriga University of Applied Sciences in Berlin where he built up the MBA program Communication & Leadership. He is author of several books, studies and essays dealing with corporate communications as well as online communications and social media. Furthermore, he has been chairman of the jury of the Pan-European Digital Communication Awards honoring excellence in professional online communications and regularly works as reviewer. In addition to his academic career he draws on vast experience as a practitioner and consultant in corporate and brand communications.



Rebecca **Böckenholt** (*Zimmermann), M.Sc. born 1989 in Bochum/Germany currently works at the International School of Management (ISM) in Dortmund as associate to the managing director. Before moving to this position in 2019, she was a research associate at the Brand & Retail Management Institute @ ISM, a university's own institute at ISM. She is responsible for the conception and conduction of quantitative and qualitative studies, in particular in the areas of retail, e-commerce and consumer behavior. Before her time at ISM, she worked as a project assistant in a medium-sized management consultancy in Bochum. She completed her studies in economics with a focus on operations research and quantitative methods/statistics at Ruhr-Universität Bochum with a Master of Science degree in 2013.



Mijka **Ghorbani**, born 1995 in Menden/Germany, is a doctoral researcher in the Department of Marketing of the Strathclyde Business School (SBS) at the University of Strathclyde in Glasgow (UK). Her research focuses on brand personality and consumer-brand relationships in digital contexts. She received her Master of Arts degree in Strategic Marketing Management at International School of Management (ISM) in Dortmund (Germany) and her Master of Science degree in Management with Marketing at Edinburgh Napier University (UK) in 2019. While studying for her Master's degree at ISM, she worked as a student research assistant at the Brand & Retail Management Institute @ ISM.



Jons **Messedat**, born 1965 in Cologne/Germany studied architecture at the RWTH Aachen, the University of Stuttgart and at the London South Bank University. At the same time, he received his diploma in industrial design from Richard Sapper at the Kunstakademie Stuttgart. After his studies, he worked in Sir Norman Foster's offices in London and Berlin. This was followed by a teaching and research position as a research assistant at the Bauhaus University Weimar, which he completed in 2004 with a doctorate in corporate architecture. His further teaching activities took him to the USA as a visiting lecturer and to various universities in Germany and Switzerland in the field of Corporate Architecture and Interior Architecture. His work as an architect, designer and author focuses on the built identity of companies and institutions. He is author of publications on Corporate Architecture, Brand Spaces, Corporate Museums and Retail Design as well as Key Note Speaker on numerous podiums. He was on the panel for built Corporate Design in the Berlin Humboldt Forum and in 2018, he was appointed onto the jury of the state prize for architecture by the chamber of architects AKNDs.



Kamran, Qeis; Becker, Marcus

The Structure and Evolution of the Marketing Field: A Content Analysis of Five Decades of Research within the Academy of Marketing Science Journals

Abstract

This paper analyses five decades of research published in the journals of JAMS and AMS review (AMSR). The authors used a Latent Dirichlet Allocation (LDA) algorithm for unsupervised topic modeling. By conducting a content analysis (CA), the authors study the relationships between the research streams of marketing evolved during the four eras of the field's progress. A multiple correspondence analysis (MCA) was drawn to validate the mapping between the digitized keywords drawn from 1913 research articles with the topic labels and terms resulting from our LDA analysis. Finally, some future directions and the needed stream of literature is presented, whereupon a path towards reinvigorating marketing in era-V could be proposed. The authors suggest another possible solution to Hunt's call, hence instead of focusing on a re-institutionalization, wherein marketing may go backwards, it must focus on essential research themes of the future

Keywords:

Marketing history, literature review, JAMS, AMS Review, multiple correspondence analysis, content analysis, ranking, marketing evolution, re-institutionalization, Latent-Dirichlet-Allocation, Unsupervised-topic-modeling-algorithms

1 Introduction

This paper studies the evolution of the marketing literature based on an analysis of all papers published in the Journal of the Academy of Marketing Science (JAMS) and the Academy of Marketing Science Review (AMSR), which are both leading journals in the marketing domain and published by the Academy of Marketing Science. Our research aims to not only show the structure and past evolution of the content of the marketing field, but also to find future pathways for the development of the marketing discipline. The authors apply a Latent Dirichlet Allocation (LDA) algorithm on the digitized keywords of 1913 research articles. Since LDA is an unsupervised learning algorithm for topic modeling, the authors expect to receive a less biased selection of research topics which play a key role in the last decades of marketing research. The authors take Hunt's (2020) call on the essential need of a re-institutionalization of marketing in the contemporary era. The authors contribute to focus and align future research efforts by paving and constructing a bridge, whereupon solid conceptual contributions for the future of the field could be delivered, thus proposing an of possible future paths for marketing.

The research was guided by two main questions:

- 1) What are the topics of marketing literature, and how have they changed during the five decades, since the inception of the JAMS and AMSR? And
- 2) Which author(s) are the most published within the marketing literature of both journals, and what were their contributions to the evolution of the field in terms of citations?

For answering the first question, the authors analyzed the content of the research streams and their developments over time. The authors took the grouped topics of our LDA model and put them into context with the published research articles grouped over years. By performing a content analysis (CA) with two major factors, the authors extracted a relationship-diagram of topic over years (see Figure 4). CA can be viewed as another unsupervised learning algorithm to reduce large dimensional textual data. The authors performed multiple robustness checks, qualitative assessment as well as quantitative assessments to validate the projections of our learning model. In order to answer the second question, the number of articles published per author and the respectively generated number of citations these articles received were identified. The most influential papers measured by citations per year are then related to the individual research themes of marketing.

While Jones and Monieson's (1990) research examined a thorough analysis of the history of marketing thought in JAMS, an updated historical account is still necessary from a contemporaneous view so that the field can reflect upon its history and propose the future steps of going forward. A keyword search revealed that since then, no other article published in JAMS or AMSR included a quantitative literature review of the marketing field. Hence, the special issue of 48 years of JAMS and 10 years of AMSR was a solid opportunity for the field to reflect upon the challenges marketing is facing today and to establish a revised historical account since the 1970s up until now at least from the perspective of the AMS.

Established journals such as JAMS and AMSR are influential in shaping the ongoing theoretical and empirical research by focusing on specific topics and reviewing submitted papers according to their set frameworks. Therefore, an analysis of the current evolution in research presented in these journals and the causes and drivers of this evolution is important. The gained insights may also help to improve marketing's position in the research literature of the future, as new challenges and opportunities for the field can become visible and prescribe solutions for vital societal challenges.

The paper is structured into four main parts. First, an introduction to the development of marketing is provided and the methods of our statistical analysis are described in detail. Second, the results of the analyses are presented, starting with a factorization of the articles into meaningful topics. Further, the authors identify trends of the discipline by conducting a time series analysis. Next, the most cited papers and most influential authors in the marketing field are assessed before and finally, a conclusion opens a discussion about the future of the marketing literature and future developments of the field.

In the analysis, the authors focus on the journals JAMS and AMSR as those are the most influential journals in the field of marketing science. Therefore, journals in the broad field of marketing as the journal of Marketing are neglected.

2 Overview of Traditional Literature Review in the Marketing Field based on JAMS and AMSR

To reflect on the possible future developments of marketing research, a brief historical account is necessary. The authors produce a *broad* but *non-exhaustive* historical account *avant la lettre* to stimulate the debate on the history of the field. While many solid historical accounts have illustrated and guided the intellectual developments of the field (Bartels 1962, 1967, 1988; Sheth et al. 1988; Wilkie/Moore 2003b; Alderson et al. 2006; Shaw and Jones 2005; Shaw 2009; Tamilia 2011; Vargo/Lusch 2004, 2008a, 2008b; Vargo/Morgan 2005; Vargo et al. 2006; Powers 2012; Kotler and Levy 1969; Levy/Kotler 1979; Kotler 1972, 2005; Luck 1969; Hunt 2020; El-Ansary et al. 2018), our historical account here is constructed on contributions and developments that compliment these works. Our brief but preliminary general overview of marketing history is established based on Hunt (2020): "For-re-institutionalizing the marketing discipline in Era V"; Wilkie and Moore (2003): "Scholarly Research in Marketing: Exploring the "4 Eras" of Thought Development; and Kotler (2005): "The Role Played by Broadening of Marketing Movement Played in the History of Marketing", and El-Ansary et al. (2018): "Marketing's identity crisis: insights from the history of marketing thought". While the authors find Key et al.'s (2020) analysis on the state of unsettledness of marketing, Yadav (2010) on the state of conceptual papers in marketing journals, and Yadav (2020) providing a contemporary view of the state of top doctoral programs in marketing in the US to be highly relevant for contemporary marketing scholars.

The essential challenge that the authors see, however, is not the general consent on the historical narrative, but rather to propose a future direction for moving forward with our field while reviewing the past. Our analysis derives from the essential paths the published research took in JAMS and AMSR so far.

Research has identified that the challenges that marketing is facing contemporarily and in the future are highly disruptive, hence the authors need to act and prepare for them today, while still coping with the fractured state of marketing of the present (Kamran 2021; Kamran et al. 2020; Key et al. 2020).

The last five decades were pivotal in the development of marketing theory and practice, with significant advances in marketing science and the development of rigorous methodologies for the field (cf. Sheth 2021). Marketing as a discipline, with a clear agenda of an autopoietic body of knowledge, was founded based on three dimensions that took shape historically (Sheth et al. 1988):

- 1) Developments of channels were ignored by economics, leading to the institutional school of marketing.
- 2) Product developments leading to the commodity school of marketing, and
- 3) Concentration on functions or processes, leading to the functional school of marketing.

However, to give an inclusive account of the history the authors need to add the contributions of the service school of marketing as a possible 4th dimension deriving from a combination of research streams, where a broader horizon of foresight for the field was needed beyond the packaged goods industry, wherein marketing was originated. Furthermore, the service school delivered a rich ground for the reproduction of knowledge within marketing theory and its practice (see Figure 5):

- 4) Developments of service thought of marketing (Skålén et al. 2015; Vargo/Lusch 2004; Shostack 1977; Grönroos 1982, 2006; Gummesson 1979; Gummesson 2008)

Era IV of marketing (Wilkie/Moore 2003; Hunt 2020) had to cope with many changes such as the globalization of markets (Levitt 1983) and many technological changes in Boorstinian's term of "*the Republic of Technology*" (Boorstin 1978) to which Neill Postman referred to as "Technopoly" (Potsman 1992), while the foundational developments have actually taken place within era III. Therefore, the dimensions that marketing developed in that era are still valid today. These foundations are:

- Brand image (Gardner/Levy 1955)
- Market segmentation (Smith 1956)
- Marketing management concept (Mckitterick 1957)
- The 4 P's of marketing (McCarthy 1960)
- Marketing myopia (Levitt 1960)
- The marketing mix (Borden 1964)
- Marketing management (Kotler 1967)

Shaw and Jones (2005) identify many schools of thought in marketing, starting from Plato and Aristotle (Shaw 1995) to Grether (1976). The schools of thought that were identified are:

- Marketing functions school
- Marketing commodities
- Marketing institutions
- Marketing management
- Marketing systems
- Consumer behavior
- Macro marketing
- Exchange, and
- Marketing history

Hult (2015) has given a solid overview of the literature and intellectual structure of marketing in JAMS (2010-2015), clustering the essential themes and topics and illustrating the essential authors. Yadav (2010) has also established a timely analysis of 30 years (1978-2007) of publications in major marketing journals on the state of conceptual articles within the field.

El-Ansary et al. (2018) have constructed on Bartels (1974) by establishing a solid historical analysis of the three paradigms in marketing while adding the fourth dimension of market systems into the marketing field:

- The traditional paradigm (1910s to late 1950s)
- The macro to micro paradigm shift, management paradigm (late 1950s to late 1960s)
- The broadening/genetic boundary expansion paradigm (late 1960s to present)
- Market systems (Wiener 1948; Bertalanffy 1969; Ashby 1956; Beer 1959, 1971; Vargo et al. 2017).

However, the authors believe that traditional literature analysis as given are limited and cannot offer a holistic approach to combine all historical and current data. Therefore, focusing on advanced master and doctor students, an innovative approach needs to be created to receive a holistic analysis, while being time efficient.

3 Method

The proposed methodology of the research embraces many essential facets and necessary dimensions to establish itself as a contribution for future researchers, who either must cope with a large amount of data or by choice, who would like to deliver solid contributions of establishing a solid and exhaustive account of the history of the different fields and disciplines. Our main focus was to avoid bias in conducting our research, therefore all the topics that were analyzed by us went through a thorough machine learning algorithms (MLA) analysis for establishing the historic and in-depth CA. Usually, conceptual articles are not very much free from some conceptual bias that goes with these types of papers. Above all, the authors have analyzed the total amount of textual data of both journals with the help of unsupervised MLA to establish an objective account of the historical narrative. Thus, the authors have combined the scholarly analyzed "*lay of the land*" in marketing with the techniques and tools of AI for more precision and objectivity in establishing unbiased labeling of the "*topic-keyword relations*". The LDA assumes that each topic is associated with multiple terms and vice versa terms can belong to multiple topics but with a different impact on the overall topic distribution. Based on the contemporary developments in ML capacity, the authors think our methodology delivers a robust foundation not only for this research but also for additional research wherein a vast amount of textual data could be analyzed for a more holistic grounded theory design. The authors deliver a dynamic approach to highlight the evaluation of topics over time.

A CA was first performed of all papers published in the JAMS and AMSR, that are not editorial notes, errata, or commentaries, to examine the content of the marketing literature, trace its evolution and identify its main streams or subfields. This analysis is used for an objective, systematic and quantitative consideration of the published articles while allowing for an interpretation of the shifting priorities of editors, reviewers, and authors, which shaped the evolution of the field.

In this paper, a foundation of synergies between data science (advanced analytics) and marketing sciences is pursued. Hence, the authors emphasize on the further evolution of marketing within the domain of digital marketing and machine learning. Based on the traditional marketing perspective, the authors need to embrace data science and AI capabilities in the age of modern management. Furthermore, based on Alderson (1957) notion of a valid definition of marketing, the authors need to integrate advanced digital marketing capabilities to not be constantly overrun by developments within the technological domain (e.g., OpenAI, Metaverse).

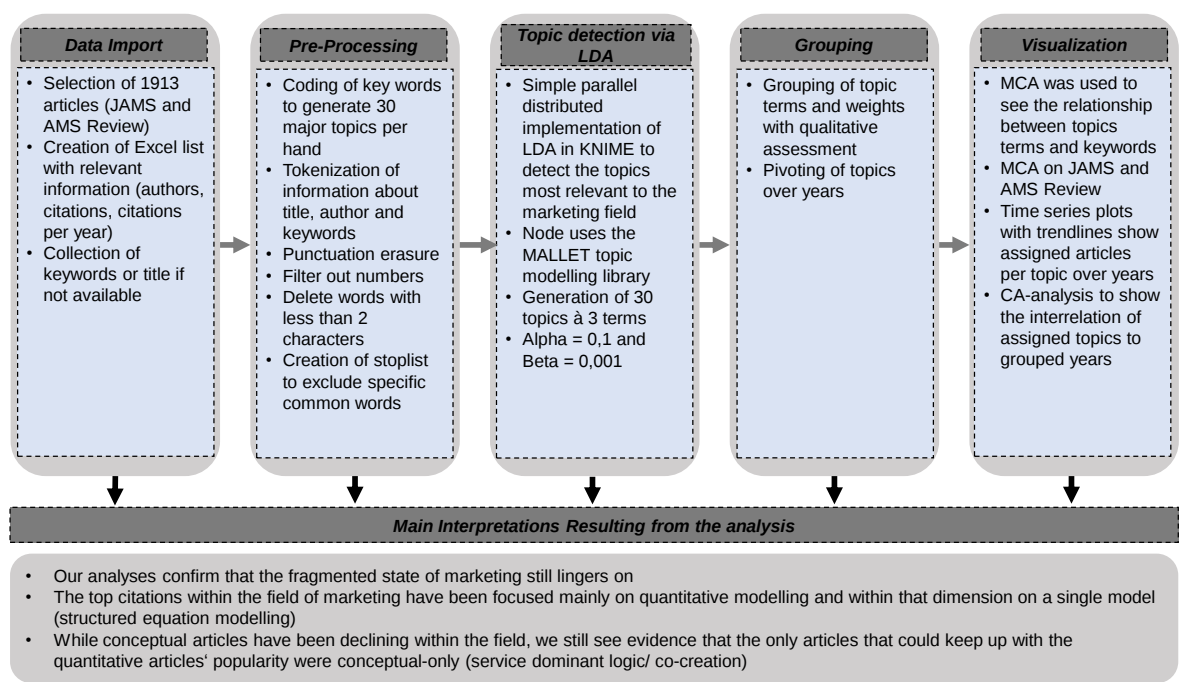


Figure 1: *Methodology of the analysis*
Source: *own illustration*

As a first step, the articles to be analyzed were selected. The authors included every article published in the JAMS and AMSR from the respective foundation of the journal until our analysis began in February 2021 (Volume 49, Issue number 2 of JAMS; Volume 10, Issue number 3 of AMSR), except for articles written by the editors, errata, commentaries, and introductions to special issues. A total of 1913 articles were selected and analyzed. The number of articles per journal and per year are shown in Table 1. JAMS has a share of 91.8% and AMSR represents 8.2% of all articles analyzed. For our analysis, the authors collected all keywords, and where there were no keywords (especially in the early issues of JAMS no keywords were indicated), the authors used the titles of the papers to analyze their content.

Table 1: *Output per journal per year*

Year	AMS Rev	JAMS	Total
1973		20	20
1974		43	43
1975		36	36
1976		37	37
1977		40	40
1978		30	30
1979		33	33
1980		33	33
1981		34	34
1982		33	33
1983		34	34
1984		40	40
1985		52	52
1986		31	31
1987		38	38
1988		35	35
1989		34	34
1990		34	34
1991		37	37
1992		39	39
1993		36	36
1994		31	31
1995		31	31
1996		26	26
1999		12	12
2000		37	37
2001		21	21
2002		33	33
2003		25	25
2004		27	27
2005		38	38
2006		48	48
2007		46	46
2008		43	43
2009		38	38
2010		50	50
2011	13	51	64
2012	8	48	56
2013	20	40	60
2014	6	38	44
2015	11	43	54
2016	14	43	57
2017	14	44	58
2018	19	56	75
2019	25	58	83
2020	27	60	87
2021		20	20
Total	157	1756	1913

Source: *own table*

3.1 Pre-processing

For conducting our analyses, the authors used the machine learning software “KNIME”, which is an API commercial software with an AI drag and drop solution. 30 major topics were identified by coding the keywords of the articles. This list of main keywords or themes was necessary because a total of almost 4000 keywords were used by the authors of the articles during the 48 years of publications. The keywords were retrieved from the databases *Web of Science* and *CrossRef*, which were accessed through the journal homepage of the *Springer Verlag*. The information about the title, author, and keywords were then transformed into a document data type and adjusted using a word tokenizer (*OpenNLP English WordTokenizer*). Tokenization is the process of chopping the given sentence into smaller parts (tokens), which is used in tasks such as spell-checking, processing searches, identifying parts of speech, sentence detection, and document classification (TutorialsPoint 2021). The raw text is tokenized based on a set of delimiters, e.g., whitespaces and punctuation. Numbers are filtered out and words with two or fewer letters are deleted. Next, a built-in “stop list” which is integrated into KNIME and contains specific words that are very common in the English language (see Appendix 6), and a manual one (see Appendix 4) with additional very common words in marketing that the authors detected were excluded in the analysis because their inclusion would distort the results with their weight and impact.

3.2 Topic Detection via LDA

A simple parallel distributed implementation of LDA following Newman et al. (2009) with the Sparse LDA sampling scheme and data structure from Yao et al. (2009) was used to detect the topics most relevant to the marketing field during the 48 years of literature analyzed. Topic relevance was firstly assessed by the topic weightings given as output measure by the LDA algorithm and is subsequent quantitative assessment judged by proximity to the originally published article keywords to the topic-terms distributions (see the following MCA) analysis. The authors further embedded the topics into a dynamic context in the sense of a per-decade and topic assessment induced by an additional CA-Analysis. The LDA algorithm uses the “*MALLET: A Machine Learning for Language Toolkit*” topic modeling library. MALLET is a Java-based package for statistical natural language processing, document classification, clustering, topic modeling, information extraction, and other machine learning applications to text (cf. McCallum 2002).

Probabilistic topic modeling is a branch of unsupervised Machine Learning algorithms that aim to annotate large archives of documents with thematic information (Blei 2012). As the model is unsupervised it does not require any a-priori labeling of the to be analyzed documents. The topics emerge from the analysis of the original texts in form of a dimension reduction similar to the numeric counterpart method of a PCA. The only difference is that LDA aims to reduce the information of textual (therefore-non-metric) data with the use of Bayesian statistics. With respect to previous topic modeling

algorithms such as the prior developed probabilistic latent thematic analysis (pLSI), LDA assumes that all documents in the collection share the same set of topics, but each document exhibits those topics in different proportions (Blei 2012). From a model-theoretic-based approach, a topic is a distribution over a fixed number of words (also referred to as vocabulary). It is assumed that the topics are to exist before any data has been generated.

LDA is a three-level hierarchical Bayesian model, in which each word of a collection is modeled as a finite mixture over a pre-existing set of topics. Each topic is, in turn, modeled as an infinite mixture over a set of topic probabilities (Blei et al. 2003). The three-level process can be characterized as follows: Each document exhibits all topics in different proportions (step 1). A randomly selected topic is chosen from the per-document distribution over topics (step 2). Then, each word in each document is drawn from the previously chosen topic (step 3). In order to base our model on solid methodological ground, the authors would like to closely describe the probabilistic model in more detail, based on the preceding literature such as Blei (2012). The underlying problem traces back to the calculation of the so-called posterior distribution, i.e., the conditional distribution of the hidden variables for the given documents. To highlight these hidden variables, let $\beta_{1:K}$ denote the topics, where each β_k is a distribution over words. The topic proportions for the d -th document are θ_d , where $\theta_{d,k}$ is the topic proportion for topic k in document d . The topic annotations for the d -th documents are z_d , where $z_{d,n}$ is the topic annotation for the n -th word in document d . The observed words for document d are w_d , where $w_{d,n}$ is the n -th word in document d , which is an item from the fixed vocabulary that stems from the automatic read-out bag-of-words representation. With these notations, the authors are able to describe the generative nature of the LDA process by using basic stochastic calculus on conditional probability. With this, the joined distribution of hidden and observed variables can be represented by

$$p(\beta_{1:K}, \theta_{1:D}, z_{1:D}, w_{1:D}) = \prod_{i=1}^K p(\beta_i) \prod_{d=1}^D p(\theta_d) \left(\prod_{n=1}^K p(z_{d,n} | \theta_d) p(w_{d,n} | \beta_{1:K}, z_{d,n}) \right)$$

The above formula uncovers various dependencies. First, the topic assignment $z_{d,n}$ depends on the per-document topic proportions θ_d . Second, the observed word $w_{d,n}$ depends on the topic assignment $z_{d,n}$ as well as on *all* topics $\beta_{1:K}$. These dependencies define LDA, which are further summarized in the graphical model of figure 2. In this figure, each node displays a random variable according to its role in the generative process. Shaded nodes characterize the words of the documents. The rectangles denote replication. The N plate denotes the collection of words within documents; the D plate denotes the collection of documents within the collection (Blei et al. 2003).

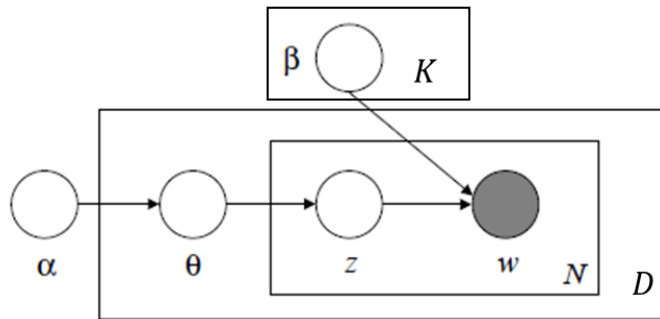


Figure 2: The graphical model for LDA according to Blei et al. (2003). Hidden variables are displayed by unshaded nodes whereas observable variables are depicted by shaded nodes. Rectangular boxes indicate multiple iterations.

Source: own illustration

Finally, the posterior probabilities can be calculated as follows:

$$p(\beta_{1:K}, \theta_{1:D}, z_{1:D} | w_{1:D}) = \frac{p(\beta_{1:K}, \theta_{1:D}, z_{1:D}, w_{1:D})}{p(w_{1:D})}$$

The numerator describes the joint variable of all the random variables (topic distributions as well as document-specific topic proportion, topic assignment, and word collection). The denominator is the marginal probability of the words, which is *theoretically* computed by aggregating the joint distribution over every possible instance of the hidden topic structure. *Practically*, this distribution is hard to compute (Dickey 1983). Therefore, Attias et al. (2000) propose a variational Bayesian framework as an approximation method for the posterior probabilities assuming a Dirichlet distribution for each topic β_k . Further, Blei et al. (2003) propose to use another Dirichlet approximation to estimate the per-document topic proportions θ_d based on a fixed parameter called α (see Figure 2).

In our analysis, the authors chose the constant parameters $\alpha = 0,1$ and $\beta = \beta_k = 0,001$ as well as $K = 30$. Each topic is associated with $T = 3$ major terms, in the sense of keywords indicating the highest probability that terms are associated with that topic. As introduced above, the alpha parameter defines the Dirichlet prior on the per-document topic distributions and the prior weight of the topics in a document. The Knime library proposes to use α of normally less than 1 for all topics to prefer sparse topic distributions, i.e., few topics per document. Per default, the parameter is set to 0,1 but could be changed accordingly. The β parameter defines the prior on per-topic multinomial distribution over words. The Knime library suggests using a number *much* smaller than 1 to strongly prefer sparse word distributions. The default value is set by 0,001. As an additional robustness check one could

perform a grid search optimization over the hyperparameters α , β , K and T which is computationally intense. The authors performed a smaller variation analysis (results shown in Appendix 6) which aims to give us a first “glimpse” of the success prospects for such a grid search optimization. For this, three academic experts were asked independently to give an a-priori expectation of a number of 30 topics in the Marketing research field that were evaluated to be relevant for the recent years. In form of a “minimal Delphi experiment” (i.e., one round iteration) the group reached a consensus for a subsample of 30 topics, that afterward were compared to the (unsupervised) topic labels provided by the LDA algorithm. The final assessment led to the conclusion not to deviate from the proposed default values within the Knime library. Note, that an increasing number of topics might have led to an even higher accuracy with the expert judgment, but the increase was relatively small (approx. 13% extra match compared to the default benchmark model using 120 topics less). From the qualitative perspective, the large increase in topic numbers did not lead to better topic distributions. Some topics were redundant with respect to a majority of keywords associated with that topic, and topic labels were hard to identify (results untabulated). Also, the visualization of such a variety of topics would be impracticable in subsequent analyses. The authors, therefore, decided to use the pre-set values in Knime ($\alpha = 0,01$, $\beta = 0,001$, $K = 30$ and $T = 3$) which the authors expect to have been approved in prior practices. In comparison to this assumption, the authors would like to highlight, that prior studies such as Furrer et al. (2008) have been limited to 26 keywords (see table 1 in Furrer et al. 2008). In our study, the authors use 3 keywords à 30 topics which is a significant improvement to the former analysis.¹

Finally, topic terms and weights were then interpreted, and each topic was named by a qualitative assessment of the authors (see Appendix 5) before they were pivoted over the years. Various robustness checks have been performed to assess the quality of the resulting topic-term distributions. The authors analyzed the association of terms to keywords and topics qualitatively for an independently chosen subset of articles, which suggests that the LDA clusters were economically sound. In the following paragraph, the authors present a multiple component analysis to quantitatively validate concordance among real published article keywords and topic-term associations. The authors used a two-component based MCA, that similarly to the metric PCA equivalent aims to project the underlying variation of the data into two main factors. Enlarging the number of components does not lead to any increases in explained variation, in terms of maximum Eigenvalue criteria that are depicted in the sensitivity analysis of appendix 1. The authors, therefore, expect that our validation analysis with two main components is sufficient to test the suitability of our LDA projections.

¹ Note that the total number of presented keywords by the LDA algorithm will be less than $3 \cdot 30 = 90$ keywords since a unique characteristic of LDA allows multiple terms to be prominent (in sense of a high probability association) in multiple topics.

3.3 Limitations of the LDA Methodology and comments

One assumption that LDA makes is the “bag of words” representation, i.e., the order of the words in the document does not matter which is problematic from a semantic point of view, especially when it comes to language generation. For uncovering the sheer structure among all documents this shortcoming does not pose a major limitation to our analysis.

Another assumption is that the order of documents does not matter. As the authors will see in the below time series analysis the interest in certain topics may vary over time, therefore it may be reasonable to estimate multiple LDAs over time. In this approach, a topic would be a sequence of distributions over words. Similarly, the authors, therefore, conducted a CA (see Figure 4) to assess the topic's relevance over decades.

A third assumption about LDA is that the number of topics is assumed known and fixed. The Bayesian nonparametric topic model provides an elegant solution (Teh et al. 2006). In this setting, the number of topics is determined by the collection during posterior inference analysis, and new documents can exhibit previously unseen topics. The latter analysis could be the content of the proceeding research. In this case, the corpus of digitized articles has to be enlarged, also accounting for other Journals than JAMS and AMS.

Table 2: Topic labels pivoted over the years

Topic\Year Category	1973-1977	1978-1981	1982-1985	1986-1989	1990-1993	1994-1999	2000-2003	2004-2007	2008-2011	2012-2015	2016-2019	2020-2021	Total
	No. %	No. %	No. %	No. %	No. %	No. %	No. %	No. %	No. %	No. %	No. %	No. %	
Advertising	18 21,4	12 14,3	15 17,9	7 8,3	6 7,1	7 8,3	2 2,4	4 4,8	2 2,4	4 4,8	5 6,0	2 2,4	182
Branding	8 10,7	4 5,3	2 2,7	0 0,0	3 4,0	2 2,7	2 2,7	10 13,3	6 8,0	13 17,3	19 25,3	6 8,0	167
Channel Management	3 4,8	3 4,8	10 16,1	2 3,2	1 1,6	3 4,8	4 6,5	5 8,1	6 9,7	12 19,4	10 16,1	3 4,8	157
Consumer Behavior	9 9,3	11 11,3	15 15,5	3 3,1	16 16,5	11 11,3	5 5,2	9 9,3	4 4,1	7 7,2	3 3,1	4 4,1	193
Consumer Experience	7 16,3	0 0,0	3 7,0	1 2,3	2 4,7	0 0,0	2 4,7	4 9,3	4 9,3	4 9,3	13 30,2	3 7,0	136
Consumer Information	9 11,8	9 11,8	11 14,5	9 11,8	9 11,8	5 6,6	5 6,6	5 6,6	6 7,9	3 3,9	4 5,3	1 1,3	175
Consumer Response	6 11,5	5 9,6	4 7,7	12 23,1	6 11,5	3 5,8	1 1,9	5 9,6	2 3,8	3 5,8	3 5,8	2 3,8	148
CSR	5 7,8	5 7,8	2 3,1	0 0,0	2 3,1	2 3,1	1 1,6	14 21,9	8 12,5	6 9,4	16 25,0	3 4,7	159
Customer Relationship	0 0,0	0 0,0	1 0,7	5 3,6	4 2,9	13 9,3	19 13,6	17 12,1	22 15,7	23 16,4	28 20,0	8 5,7	234
Customer Satisfaction	0 0,0	3 4,5	0 0,0	4 6,0	4 6,0	5 7,5	9 13,4	4 6,0	15 22,4	8 11,9	10 14,9	5 7,5	160
Firm Performance	3 2,4	3 2,4	4 3,2	6 4,8	8 6,5	5 4,0	19 15,3	15 12,1	17 13,7	22 17,7	20 16,1	2 1,6	222
Food Products	2 6,7	1 3,3	4 13,3	1 3,3	0 0,0	1 3,3	3 10,0	1 3,3	4 13,3	4 13,3	6 20,0	3 10,0	120
Healthcare	6 14,6	2 4,9	2 4,9	8 19,5	8 19,5	0 0,0	0 0,0	2 4,9	0 0,0	0 0,0	9 22,0	4 9,8	131
Industries	14 29,8	5 10,6	5 10,6	5 10,6	5 10,6	1 2,1	0 0,0	1 2,1	3 6,4	3 6,4	3 6,4	2 4,3	143
Innovation Management	3 3,8	1 1,3	2 2,5	5 6,3	5 6,3	3 3,8	3 3,8	9 11,4	15 19,0	11 13,9	13 16,5	9 11,4	168
International Marketing	2 5,1	6 15,4	6 15,4	1 2,6	5 12,8	2 5,1	2 5,1	2 5,1	4 10,3	6 15,4	2 5,1	1 2,6	136
Model and Measurement	4 9,5	0 0,0	3 7,1	7 16,7	7 16,7	3 7,1	1 2,4	0 0,0	3 7,1	8 19,0	4 9,5	2 4,8	137
Model Effects	6 16,2	1 2,7	2 5,4	6 16,2	1 2,7	0 0,0	1 2,7	4 10,8	5 13,5	4 10,8	6 16,2	1 2,7	134
Performance	1 2,0	2 4,1	0 0,0	4 8,2	5 10,2	2 4,1	8 16,3	4 8,2	4 8,2	9 18,4	8 16,3	2 4,1	145
Pricing	10 14,7	4 5,9	7 10,3	4 5,9	4 5,9	2 2,9	3 4,4	9 13,2	3 4,4	10 14,7	9 13,2	3 4,4	164
Product Development	5 13,5	0 0,0	1 2,7	5 13,5	1 2,7	0 0,0	5 13,5	3 8,1	1 2,7	7 18,9	7 18,9	2 5,4	132
Public Theory	9 12,5	3 4,2	4 5,6	4 5,6	7 9,7	5 6,9	1 1,4	2 2,8	10 13,9	7 9,7	11 15,3	9 12,5	160
Retailers	6 9,5	15 23,8	14 22,2	4 6,3	8 12,7	3 4,8	2 3,2	3 4,8	1 1,6	4 6,3	3 4,8	0 0,0	163
Sales Management	9 8,0	10 8,9	12 10,7	9 8,0	12 10,7	11 9,8	7 6,3	11 9,8	8 7,1	11 9,8	7 6,3	5 4,5	208
Scaling	5 9,1	5 9,1	4 7,3	4 7,3	6 10,9	4 7,3	4 7,3	7 12,7	5 9,1	4 7,3	4 7,3	3 5,5	150
Service Dominant Logic	2 4,9	2 4,9	1 2,4	2 4,9	0 0,0	2 4,9	3 7,3	1 2,4	14 34,1	4 9,8	9 22,0	1 2,4	139
Shopping Behavior	11 19,3	5 8,8	17 29,8	10 17,5	5 8,8	1 1,8	0 0,0	3 5,3	1 1,8	2 3,5	2 3,5	0 0,0	157
Social Media	1 1,9	0 0,0	1 1,9	0 0,0	1 1,9	0 0,0	0 0,0	3 5,8	4 7,7	8 15,4	26 50,0	8 15,4	137
Strategy	3 4,3	2 2,9	2 2,9	3 4,3	1 1,4	4 5,8	4 5,8	2 2,9	16 23,2	7 10,1	12 17,4	13 18,8	150
Surveys	9 23,1	11 28,2	5 12,8	7 17,9	4 10,3	0 0,0	0 0,0	0 0,0	2 5,1	0 0,0	1 2,6	0 0,0	139
Total	176	130	159	138	146	100	116	159	195	214	273	107	4744

Source: own table

3.4 Analyses

A multiple correspondence analysis (MCA) was used to see the relationship between the topic terms and keywords. MCA is an exploratory data analysis technique for the graphical display of multivariate categorical data (Benzecri 1982; Hoffman/Franke 1986; Lebart 1984) and aims to explain the interdependence among a set of categorical variables (keywords and topics) that is similar to a principal component analysis (Hoffman 1994). It allows the researcher to explore and analyze multi-way tables in order to detect a structure in the relationships between nominal variables (see Furrer et al. 2008). The analysis shows which rows and columns of a frequency table have similar patterns of counts.

As depicted in Figure 3 collected keywords from the articles (blue points) and associated terms (green points) and subordinate topics (orange point) overlap perfectly thus underpinning the good fit of the LDA topic extractor. In this analysis, the authors chose a two-dimensional factor, which is the most suitable representation for a graphical outline. The corresponding eigenvalues were both around one, with a total inertia of about 655 and explained inertia of 0.15% for each component. As noticed by Kaciak and Louviere (1990), the proportion of the total variance explained by the dimensions is often very small. This is closely connected to the (inevitably) binary nature of the transformed nominal data (Lebart 1984). Additional sensitivity analyses for higher orders of components ($n = 2-30$) have been performed, which did not lead to an improvement of the maximum explained inertia per component (see Appendix 1).

Based on our observations within the research and our experience as marketing community the authors would like to propose another possible solution to Hunt's (2020) call, thus, instead of focusing on a re-institutionalizational framework for marketing, that marketing science does not only look back at what has been lost but simultaneously for what is to come.

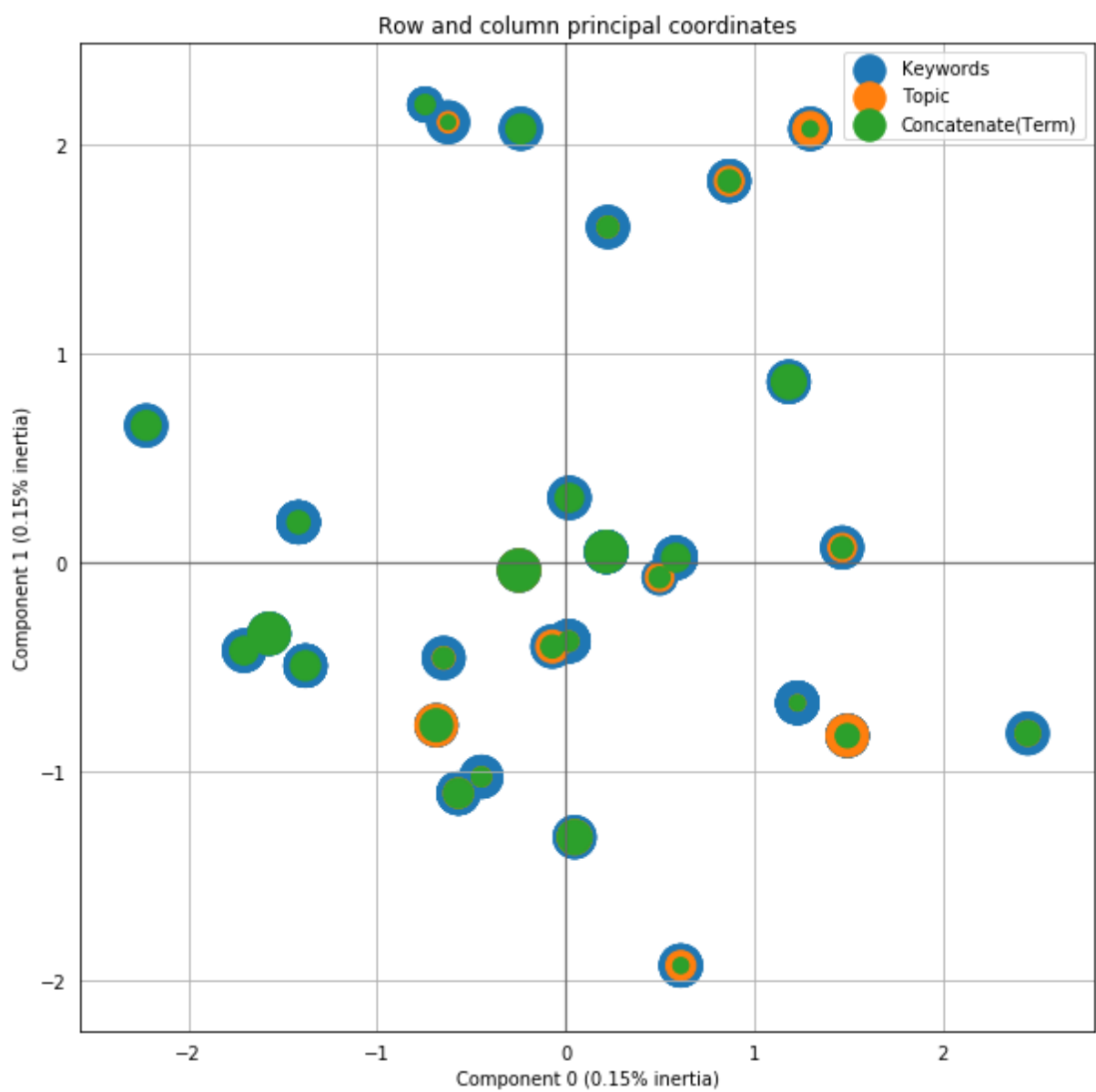


Figure 3: MCA Analysis showing the proximity of observable research article keywords (blue dots), topic assignment (orange dots), and (concatenated) terms assignment (green dots). Dot size indicates the number of articles affiliated with the categories.

Source: own illustration

The MCAs on the JAMS (1973-2021) and the AMSR (2011-2021) articles can be seen in appendices.

Appendix 2 and Appendix 3 showing equally good results. In principle, the above MCA results indicate a high match between the observed literature articles and the topic-term distributions. In succeeding analysis researchers can assess advanced visualization methods as presented in the Python library LDA-vis developed by Sievert et al. (2014). Moreover, the authors define a new *relevance* score to assess the affiliation of a term to a topic by a mixture of the log-probability of the term $w_{d,n}$ and the approximated Variation Bayesian distribution (as introduced in the theory part above) over the empirical term frequency. This measure overcomes shortcomings of simply ranking terms purely by their probability under a topic as suggested by prior studies of Bischof and Airoldi (2012). The authors performed a comparable visualization method to Sievert et al. (2014) by using an MCA among topic-term distributions as described in the previous section. In order to also highlight the dynamic changes of topics over time, the authors further perform a CA between topic and years association.

3.5 CA-Analysis/Results

To show the interrelation of the assigned topics to grouped years, a CA-Analysis was carried out (see Figure 4). The frequency of published articles within a topic was based on a pivoting table drawn from the assigned topics over years after LDA extraction. Again, the authors chose a two-factor model to simplify the graphical representation. Topics (blue bubbles), that are closely related to a time period (orange bubbles), are also geometrically (in terms of the Euclidean distance) close together. The bigger the topic bubble, the more articles were published in the topic, and the bigger the years' bubble, the more articles were published in the respective timespan. The analysis provides a pair of coordinates in the two-dimensional space for each of the 1913 articles included in the dataset. As such numerous dots (for each article) in the map would become uninterpretable, the authors decided to only represent the position of the topics. However, as the previously discussed MCA analysis suggests, article keywords and topics are closely correlated to each other (see Figure 3, as well as the per-journal, conducted MCA in appendices 2-3).

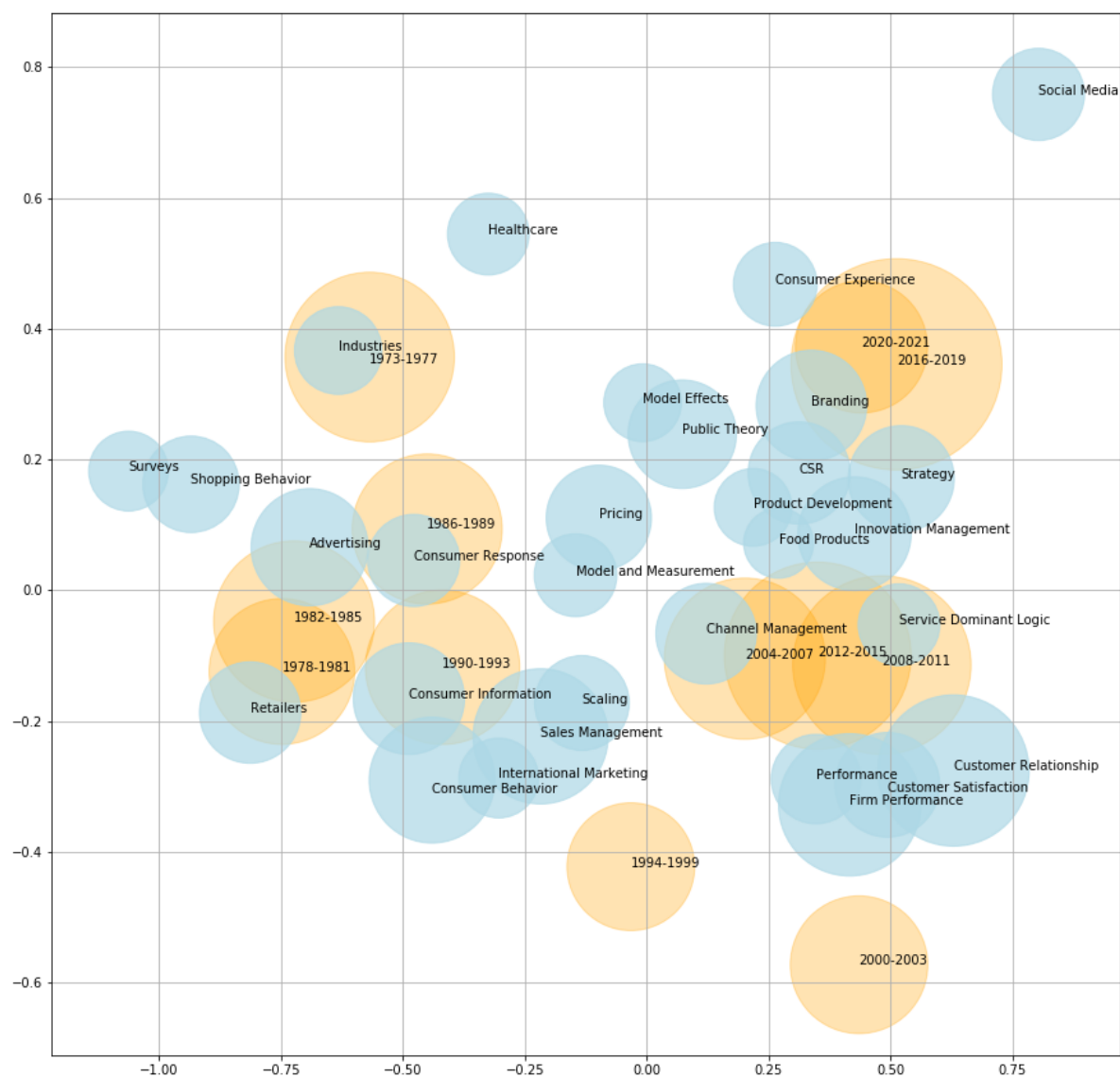


Figure 4: CA-Analysis indicating the association of topics (blue bubbles) with time decades (orange bubbles). Size of the bubbles indicates the number of associated articles to either topic and year category respectively. Bubbles closer together indicate thematic proximity induced by the underlying LDA classifier.

Source: own illustration

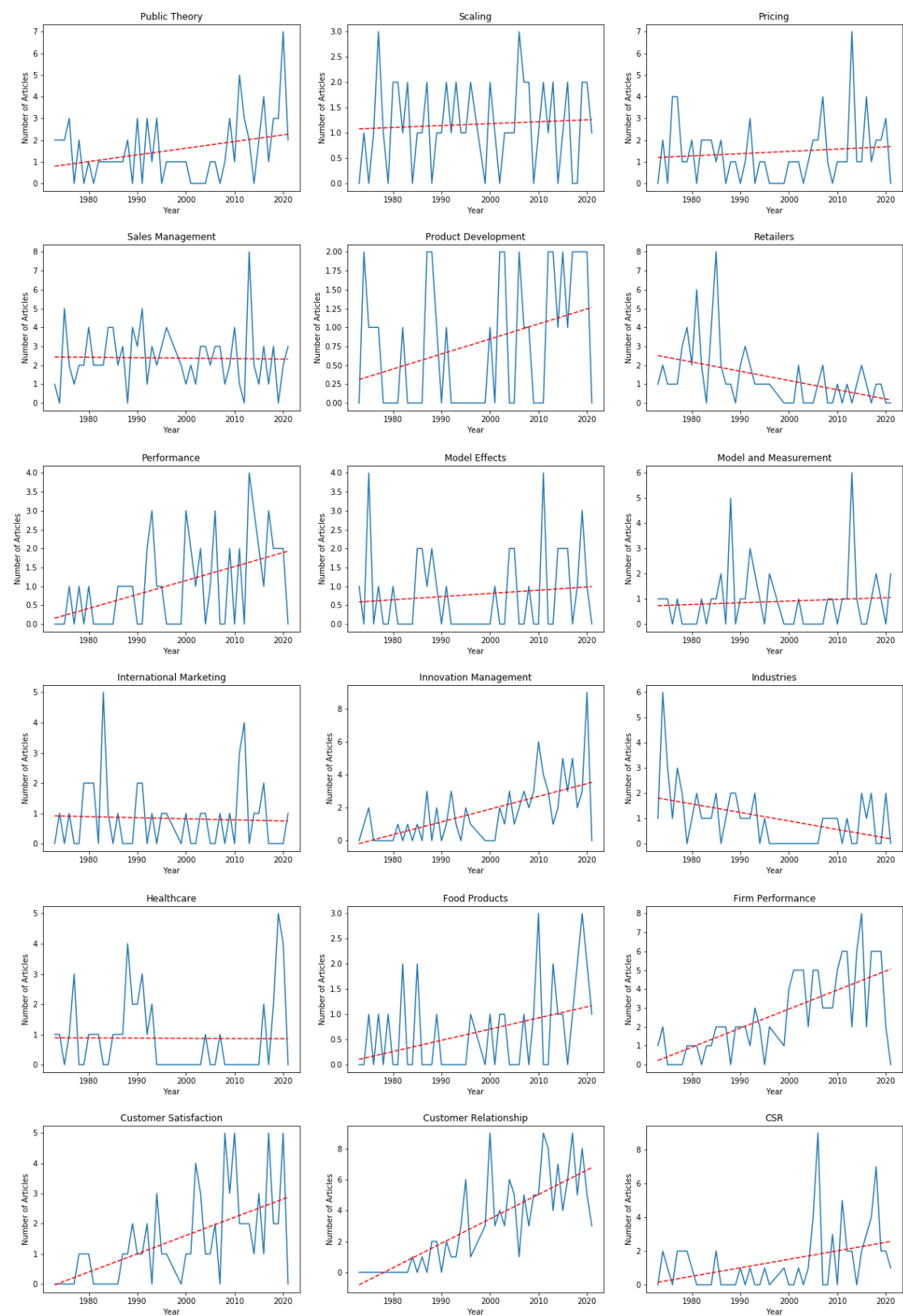
The two dimensions of the map in Figure 4 can be interpreted as follows. The horizontal dimension separates older topics (on the left) from newer ones (on the right), with a cut around the year 1999. Whereas the vertical axis separates topics focusing from a qualitative assessment with a focus on consumer individual aspects (at the top) and company-related aspects (at the bottom). These dimensions reflect “poles” of topical orientation within marketing. The combination of the two dimensions provides insights into the relationships between the topics.

The timeline follows a spiral form. On the top, left are older articles with topics such as “Industries”. Below that are articles from the 90s, when topics as “Scaling”, “Sales Management”, “Consumer Behavior” and “International Management” were important. The turn of the millennium is represented in the bottom right quadrant, when especially topics as “Customer Relationship” and “Customer Satisfaction” became more significant. This trend continues until today. At the top right the authors find the newest articles. There is a clear trend towards social media (top right), but also other topics such as “Corporate Social Responsibility”, “Branding”, and “Product” are strongly associated with the modern era (2016-2021). It is also worth pointing out that the authors identify so-called “evergreen” topics affiliated to all decades, therefore placed in the center of Figure 4 such as “Model and Measurement”, “Pricing”, “Model Effects” and “Public Theory”. But also, traditional topics, as “Healthcare” have been found increasing awareness to modern ages, i.e., in geometrically terms have been pushed to the right in the topic coordinates of figure 4, due to the ongoing Corona pandemic. Moreover, the authors notice the exposed role of “Social Media” to the far upper right of the most recent articles in 2020-2021. From our perspective, this topic should therefore receive further recognition in future marketing research since it is “detached” from the current literature clusters. To further validate our findings, the authors would like to highlight the good fit of topics like “CSR” and “Service-Dominant” logic. CSR was placed in between articles from decade 2004-2015 and 2016-2021 where the topic gained increasing notice not only in marketing literature but also in mainstream media. Same for the topic “Service-Dominant Logic” a specialization which is connected to well-perceived articles published in 2008 such as Vargo and Lush (2008) or Payne et al. (2008).

Figure 4 illustrates that using LDA for the CA approach results in a holistic analysis containing all historical and current data. This systematic model leads as well to a time efficient literature analysis.

In order to assess the evolution of topics over time, the authors performed a time series analysis (see Figure 5). A positive (negative) slope of the trend line (dashed red line) indicates an increase (decrease) of the average number of associated articles to the respective topic over the years 1973-2021. To name a few major findings, the authors observe a strong increase in topics like “Customer Relationship”, “Customer Satisfaction”, and “Innovation Management”, but also “Firm Performance” still has a positive trend showing that the traditional company-driven forces have not been completely erased by the above-mentioned consumer individual aspects, whereas “old-fashioned” topics such as “Surveys”,

“Retailers” and “Advertising” have declined significantly. Surprisingly, “Consumer Information” also visibly loses awareness over time, which (by following Figure 5) was more prominent in the 90s. During the last three years (2019-2021) the authors see an obvious increase in associated citations for topics as “Social Media”, “Innovation Management” (again), “Consumer Experience”, “Strategy” as well as in “Public Theory” and “Healthcare”, whereas the latter topics’ increases are most likely connected to the ongoing Covid-19 pandemic. As already stated in the above CA, the time series analysis again reveals the notion that the “Service-Dominant Logic” only shows a short periodic prominence around 2008 after Vargo and Lusch (2008a) published their theory, supporting the validity of our LDA model. Further, the authors detect an increase in CSR literature in the mid-2000s referring to frequently cited articles such as Maignan and Ferrell (2004), Sen (2006), and Ellen (2006) proving further the robustness of the underlying LDA classifier. Again, it is worth pointing out that the underlying topic modeling methodology belongs to the category of unsupervised learning algorithms that were not fed by a-priori labeled topic or terms names as in prior literature studies such as Furrer et al. (2008). The algorithm simply combined thematic close keywords into groups, that as the prior robustness checks highlight seem to be meaningful from an expert marketing perspective. It, therefore, overcomes subjective topic labeling by researchers that are embedded into a certain research environment, which might yield biased topic assessments. The only qualitative assessment, that our research team had to deliver was the sub-summation of terms to topics which in many settings like for terms such as “Corporate”, “Social” and “Responsibility” to CSR was self-explanatory.



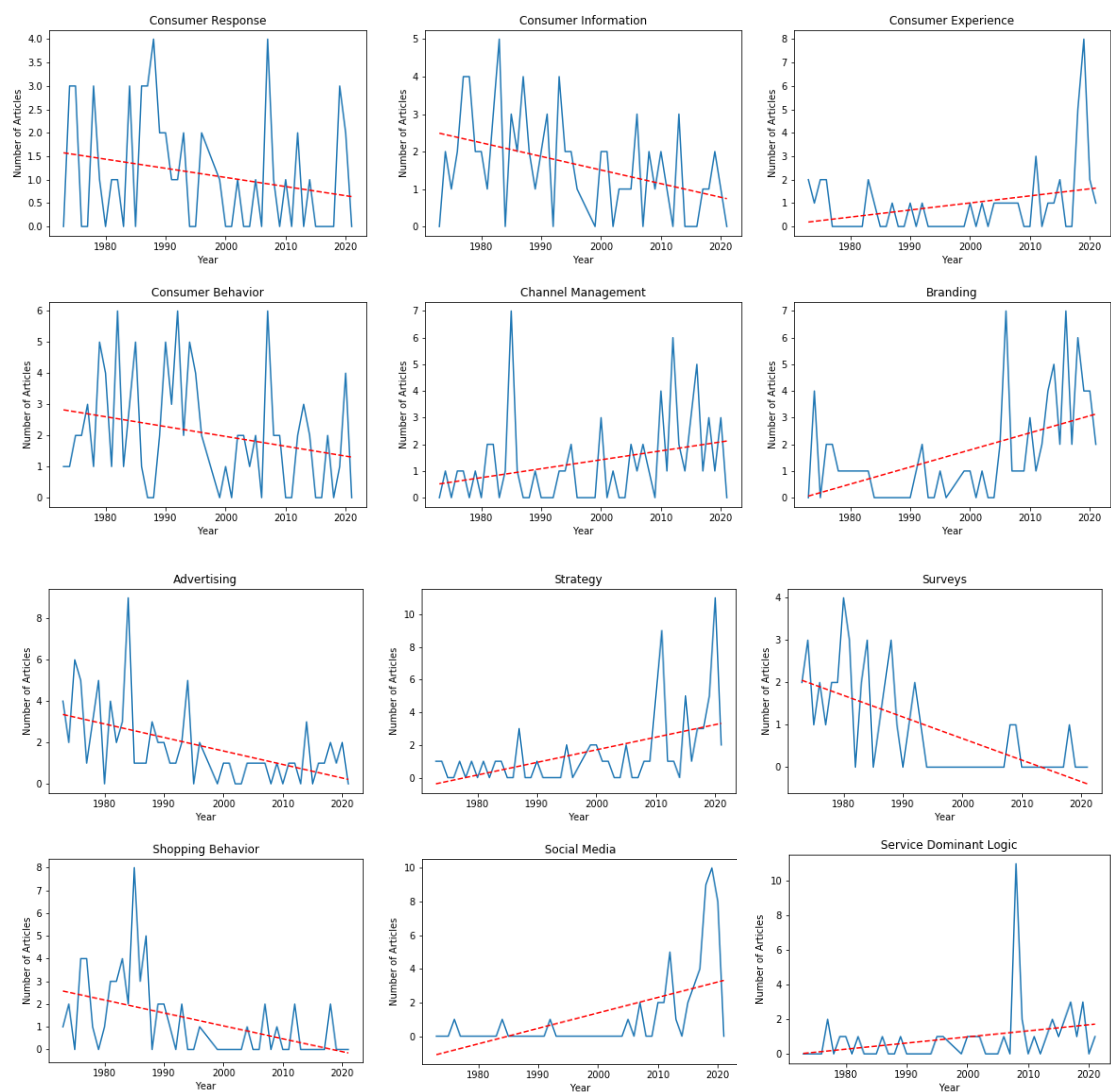


Figure 5: Time Series Plots (blue) with trendlines (red) showing the evolution of articles per topic over time
Source: own illustration

Aggregating over the years underlines the relevance of the top 3 topics “Customer Relationship”, “Firm Performance” and “Sales Management” that have an increasing or at least constantly high level of publications. Fewer detections can be found in “Food Products”, “Product Development”, “Surveys” as well as in “Model Effects” whereas the authors see a clear increase in labeled “Food Product” articles throughout the recent years underpinning an ongoing interest in nutrition of the general population (Barnard 2019).

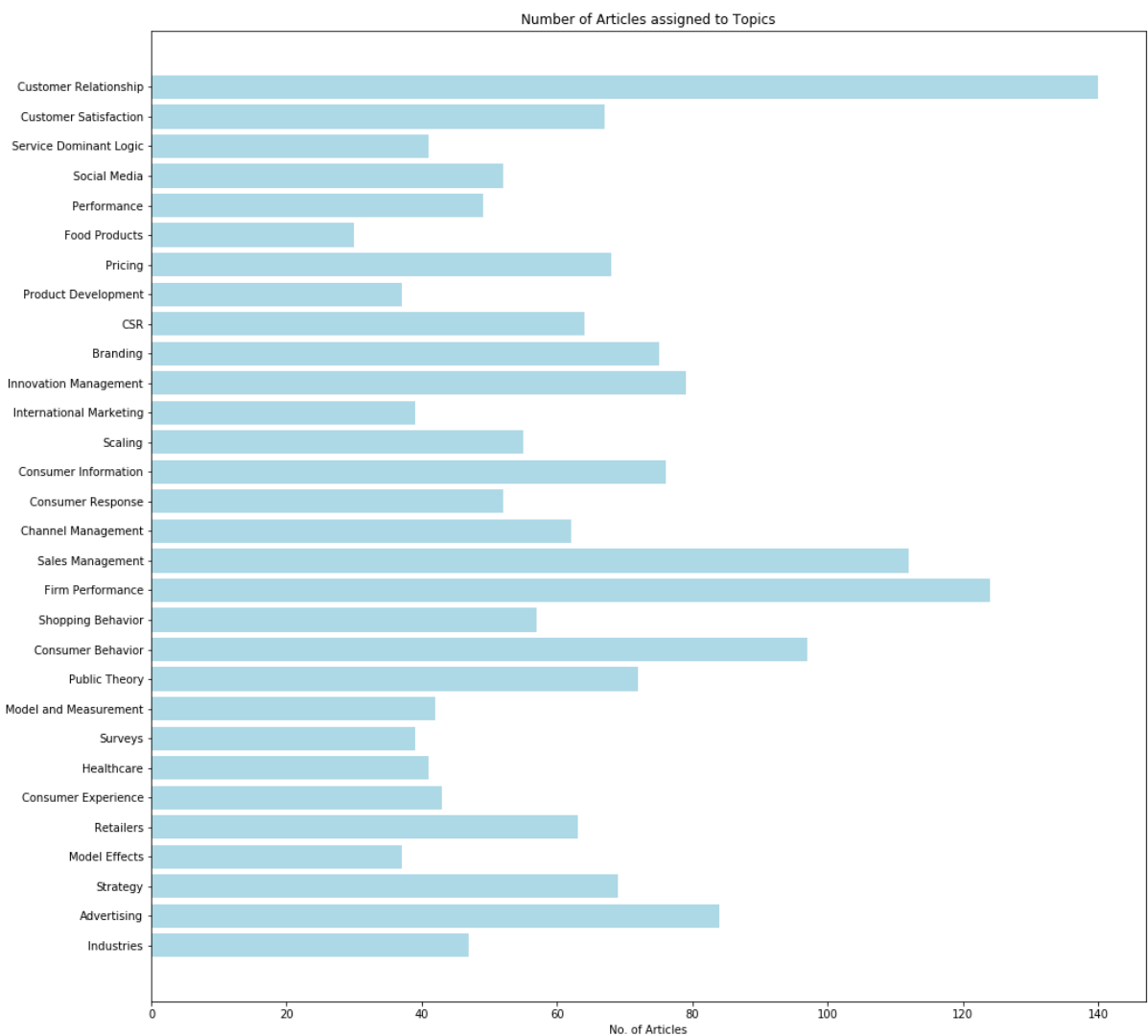


Figure 6: Number of articles assigned to topics
Source: own illustration

Investigating the evolution of the total marketing literature in JAMS and AMSR over time, the authors see a clear increase from the millennium to 2019 whereas the drop in the last period stems from the fact that the authors collected the data at the beginning of 2021 resulting in a significantly shorter period for the last time bucket in comparison to 4 years in total for the remaining time periods. Concluding that research in marketing is emerging from its relatively low at the end of the 90s to a steadily increasing publication activity over the years.

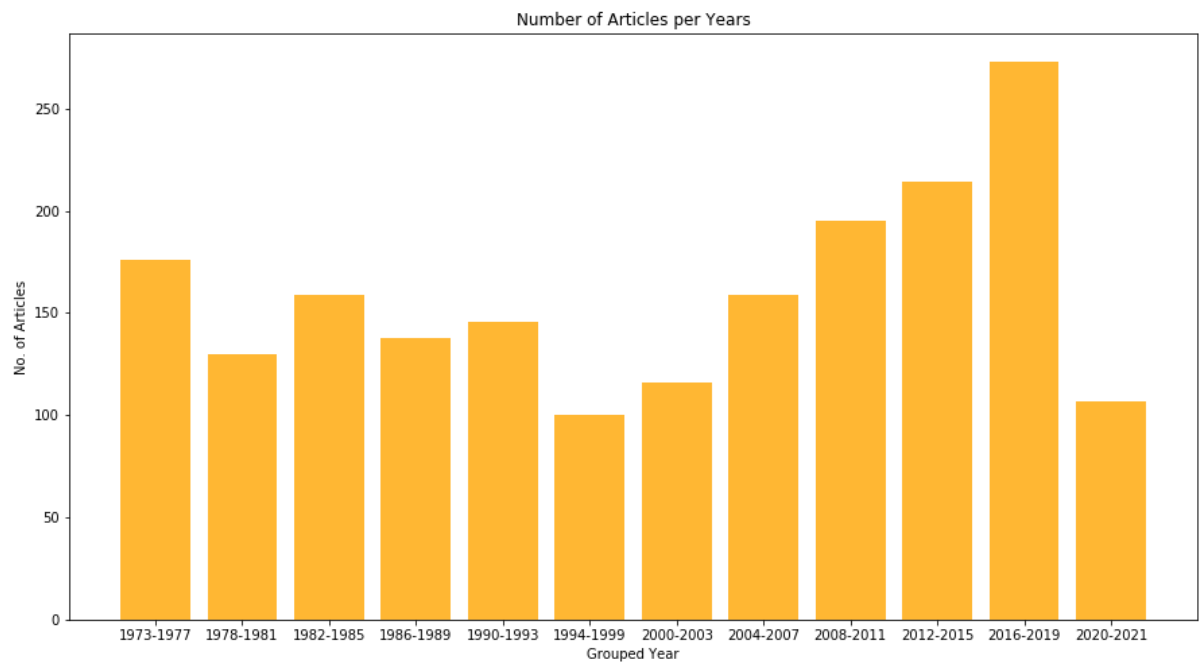


Figure 7: *Number of articles per timespan*
Source: *own illustration*

3.6 Contributing Authors

The study by Bergh et al. (2006) found that the author’s characteristics have the most explanatory power on an article’s impact. They concluded that researchers who published the most in a period had the strongest impact on the themes studied during the following period. Therefore, the authors analyzed the most published authors in marketing to better understand the past evolution of the field and possible future directions.

The authors identified and ranked those authors whose work has appeared most frequently in JAMS and AMSR. The 1913 articles have been written by more than 2300 different authors. Among these authors 78 (3.3%) published 5 articles or more, 76 (3.2%) published 4 articles, 149 (6.3%) published 3 articles, 404 (17.2%) published 2 articles, and 1643 (69.9%) published only 1 article. Thus, only 30.1% of all authors have published more than one paper in all of JAMS’ and AMSR’s history. These proportions are comparable to those found by Furrer et al. (2008) in strategic management and Heck and Cooley (1988) in the financial literature.

Table 3 shows a ranking of the top 40 authors based on both the total number of appearances and adjusted appearances, which reflect multiple-authored articles. In scientific literature analysis, adjusted appearances refer to a method of quantifying an author's contribution to research publications when multiple authors have collaborated on an article. This adjustment aims to distribute credit fairly among the co-authors based on their level of contribution.

The method works as follows:

- 1) If two writers collaborated on an article, each author is assigned half of the credit for that publication.
- 2) If three writers collaborated, each author is assigned one-third of the credit.
- 3) This pattern continues for articles with more authors, where each author's credit is divided equally among the co-authors.

The provided statistics in the passage show the distribution of papers based on the number of authors involved. For example, five hundred and fourteen (26.9%) papers were written by one author, 656 (34.3%) by two authors, 516 (27.0%) by three authors, 181 (9.5%) by four authors, 35 (1.8%) by five authors, and 11 (0.6%) by six authors or more. This information helps provide a more nuanced understanding of authorship contributions within the dataset, considering the complexities of multiple-authored articles

The five most prolific authors were: Shelby D. Hunt, with 25.2 adjusted appearances and 32 total appearances, Rajan Varadarajan, with 15.6 adjusted appearances and 21 total appearances, William O. Bearden, with 9.8 adjusted appearances and 23 total appearances, Christian Homburg, with 8.3 adjusted appearances and 22 total appearances, and Richard P. Bagozzi, with 8.3 adjusted appearances and 14 total appearances.

This analysis emphasizes the contributions of researchers over an almost 50-year period, by distinguishing between long-established and new generations of scholars. This latter group will undoubtedly and increasingly be tasked with defining the marketing field's future directions. Because an author's number of published papers is proportional to the duration of his or her career, the writers at the top of our list also have the longest careers; as a result, they have had the most effect on the structure and evolution of the marketing field.

Table 3: *Most published authors in absolute number of appearances versus adjusted appearance accounting for articles with multiple authors*

Rank	Author	Total Appearances	Adjusted Appearances
1	Hunt, Shelby D.	32	25.2
2	Varadarajan, Rajan	21	15.6
3	Bearden, William O.	23	9.8
4	Homburg, Christian	22	8.3
5	Bagozzi, Richard P.	14	8.3
6	Lusch, Robert F.	16	8.2
7	Calantone, Roger J.	19	8.1
8	Kumar, V.	19	7.8
9	Grewal, Dhruv	24	7.7
10	Ferrell, O. C.	16	6.9
11	Enrick, Norbert Lloyd	9	6.3
12	Malhotra, Naresh K.	8	6.2
13	Morgan, Neil A.	12	5.6
14	Moschis, George P.	7	5.2
15	Mahmoud, Essam	7	5.0
16	Lamb, Charles W.	10	4.6
17	Singh, Jagdip	9	4.6
18	Sivakumar, K.	10	4.5
19	Teas, R. Kenneth	6	4.5
20	Sheth, Jagdish N.	8	4.5
21	Samiee, Saeed	8	4.2
22	Mentzer, John T.	9	4.1
23	Peterson, Robert A.	8	4.1
24	Sirgy, M. Joseph	6	4.0
25	Luo, Xueming	8	4.0
26	Palmatier, Robert W.	11	3.8
27	Brown, Stephen W.	9	3.6
28	Hult, G. Tomas M.	9	3.6
29	Diamantopoulos, Adamantios	7	3.5
30	Parasuraman, A.	9	3.5
31	Dubinsky, Alan J.	7	3.5
32	Futrell, Charles M.	4	3.5
33	Monroe, Kent B.	9	3.4
34	Sen, S.	4	3.3
35	Woodside, Arch G.	6	3.3
36	Jaworski, Bernard J.	6	3.3
37	Horton, Raymond L.	4	3.3
38	Hirschman, Elizabeth C.	4	3.3
39	Hawes, Douglass K.	4	3.3
40	Hulland, John	8	3.3

Source: *own table*

3.7 Most Influential Papers

Certain publications have played pivotal roles in the growth of every scientific subject. These papers are accelerating factors in the field's growth because of their impact (Berry/Parasuraman 1993). In order to better comprehend the future possibilities of the marketing discipline, it is necessary to determine the most prominent papers published in JAMS and AMSR between 1973 and 2021. The authors utilized the commonly accepted approach of summed citation counts to identify and assess the effect or impact of research work (Bergh et al. 2006; Ramos-Rodríguez/Ruíz-Navarro 2004; Tahai/Meyer 1999).

Table 4: *Most influential papers ordered by citations per year to account for the time horizon of long-established researchers*

Rank	Authors	Year	Journal	Citations	Citations per year
1	Henseler et al.	2015	JAMS	3312	552
2	Bagozzi and Yi	1988	JAMS	10457	317
3	Hair et al.	2012	JAMS	2230	248
4	Vargo and Lusch	2008	JAMS	3200	246
5	Vargo and Lusch	2016	JAMS	1011	202
6	Bagozzi and Yi	2012	JAMS	1153	128
7	Grönroos and Voima	2013	JAMS	1024	128
8	Payne et al.	2008	JAMS	1620	125
9	Dick and Basu	1994	JAMS	2955	109
10	Hollebeek et al.	2019	JAMS	215	108
11	Pansari and Kumar	2017	JAMS	362	91
12	Davenport et al.	2020	JAMS	77	77
13	Voorhees et al.	2016	JAMS	357	71
14	Edvardsson et al.	2011	JAMS	659	66
15	Hair et al.	2017	JAMS	262	66
16	Harmeling et al.	2017	JAMS	259	65
17	Ranjan and Read	2016	JAMS	304	61
18	Appel et al.	2020	JAMS	60	60
19	Yoo et al.	2000	JAMS	1243	59
20	Zeithaml et al.	2002	JAMS	1124	59
21	Berry	1995	JAMS	1488	57
22	Sen	2006	JAMS	803	54
23	Maglio and Spohrer	2008	JAMS	673	52
24	Lam et al.	2004	JAMS	872	51
25	Szymanski and Henard	2001	JAMS	1006	50
26	Maignan and Ferrell	2004	JAMS	834	49
27	Homburg et al.	2017	JAMS	191	48
28	Diamantopoulos et al.	2012	JAMS	426	47
29	Burnham et al.	2003	JAMS	849	47
30	Ellen	2006	JAMS	705	47
31	Lusch et al.	2010	JAMS	503	46
32	Peloza and Shang	2011	JAMS	452	45
33	Dabholkar and Bagozzi	2002	JAMS	849	45
34	Zeithaml et al.	1993	JAMS	1251	45
35	Baker and Sinkula	1999	JAMS	961	44
36	Sheth et al.	2011	JAMS	435	44
37	Wilson	1995	JAMS	1102	42
38	Zeithaml	2000	JAMS	851	41
39	Nysveen	2005	JAMS	641	40
40	Etgar	2008	JAMS	513	39

Source: own table

The most significant papers published in JAMS and AMSR were determined based on the number of times they appeared on the *Web of Science* and *CrossRef* databases, accessed on the *Springer Verlag* website. Because a paper published earlier has a greater chance of receiving more citations than one published later, the articles were sorted according to the number of citations divided by the number of years they were published, thereby following the procedure displayed in Furrer et al. (2008). The most influential works are listed in Table 4 with their number of citations and citations per year.

4 Conclusions, Limitations, and Discussion

Conclusions

Our research aimed to illustrate the structure and past evolution of the content of the marketing field but also find possible future directions for the development of the marketing thought deriving from our in-depth analysis of all the publications based on JAMS and AMSR. The authors performed various unsupervised machine learning algorithms such as LDA, MCA, and CA in order to detect major topics in the marketing research literature and put our results into a dynamic perspective. In this manner, the authors expect to receive a less biased, i.e., less subjective, view on our discipline. The authors have embedded our research within the context of Hunt's (2020) call on the essential need for a re-institutionalization of marketing in the contemporary era V, as a more advanced and contemporary historical account that also embraces an analysis.

Indeed, a history of 48 years in the case of JAMS and 10 years in the case of AMSR is a good time to reflect on the maturity of the marketing discipline constructed within AMS. The authors as marketing community are well-advised to return our attention to the volume of literature evolved so far by our field and, thus designing a solid bridge and in particular paving the way for further research (Ramos-Rodríguez/Ruíz-Navarro 2004) and to answer Hunt's (2020) call.

Within contemporary technological advances, our research methodology was also a unique combination of manually scholarly research and combining it with the support of AI-driven capability (machine-learning using KNIME Analytics Platform) to embrace much larger sets of articles for more advanced and also rigorous research findings. The authors have analyzed in total a number of 1913 articles for this study. As indicated in Table 1 of this research, 157 (8.2%) articles were published in AMSR, while 1756 (91.8%) articles were collected within the publications of JAMS. The authors believe that this approach could amplify not only that the marketing field adapts to a more AI-supported review of the field's research methods and findings, but moreover to deliver possible motivations for more robust and insightful future research.

Our research was motivated by two main questions:

- 1) What are the topics of marketing literature, and how have they changed during the past 48 years? And
- 2) Which author(s) are the most published within the marketing literature of both journals, and what were their contributions to the evolution of the field?

Our theoretical contributions could be reported in the following manner:

The authors have identified the major research streams within the JAMS and AMSR during the lifetime of both journals of the academy of marketing science. The authors furthermore studied the relationships between the subfields of marketing and thus accounting for the structural evolution of the field since its inception over time until today.

Our research illustrates the following observation:

- The dimension of the industry as a theme of research has declined in importance (1973-1977).
- A Decline in importance in retailing research in marketing (earlier 1980s) was also observed.
- A solid decline has taken place within the dimension of advertisement, especially through TV as a medium of communication.
- The era of the 1990s was predominantly motivated by research in CB, consumer information, and international marketing.
- Pricing as a theme of research has kept its validity, but at the same the low profile of importance over the span of research published.
- After the turn of the century topics as performance, customer satisfaction, customer relationship, and firm performance won on importance. A solid observation here also was the importance of channel management that reappeared in the mid-2000s as a research theme. However, the authors also could confirm the importance of channels throughout the journals' history on a constant basis.
- Changes pertaining to the evolution of the field based on our research also indicate the essential timelines, wherein the rise and the decline of the individual stream of research were embedded. While the authors also illustrate the essential peaks and furthermore the rapid declines of the themes.

After the turn of the century topics as performance, customer satisfaction, customer relationship, and firm performance won on importance. A solid observation here also was the importance of channel management that reappeared in the mid-2000s as a research theme. However, the authors also could confirm the importance of channels throughout the journals' history on a constant basis.

Changes pertaining to the evolution of the field based on our research also indicate the essential time-lines, wherein the rise and the decline of the individual stream of research were embedded. While the authors also illustrate the essential peaks and rapid declines of the diverse themes.

The essential contributors for the field based on total appearances and adjusted appearances have been identified, and the top 5 list has been occupied by 1) Shelby Hunt, 2) Rajan Vardarajan, 3) William Bearden, 4) Christian Homburg, and 5) Richard Bagozzi. This results in a total of 112 appearances and 67.1 adjusted appearances.

The top 5 most essential papers published based on total citation and per year citation index are 1) Henseler et al. (2015), 2) Bagozzi and Yi (1988), 3) Hair et al. (2012), 4) Vargo and Lusch (2008), and 5) Vargo and Lusch (2016), which encompass a total of 20,210 citations and 1,565 citations per year.

On the notion of Hunt's (2020) observations of the streams of research that could stimulate the needed disciplinary indigenous foundations whereupon the objective of a re-institutionalization of the marketing discipline could occur, the authors could confirm that the most cited papers in JAMS/AMSR are based on a quantitative nature. The authors also can confirm that Hunt's (2020) call remains open, hence not only because of the nature of the particularly quantitative modeling focus of the most cited contributions (no. 1, 2, 3, and 6), but moreover, our research has also illustrated that a single quantitative/model approach has been concentrated upon namely the structural equation modeling within all the above-mentioned articles. Solid ground for hope also has been highlighted that many of the additional contributions pertaining to the most essential publications within the field (no. 4, 5, 7, 8, and 10) have been of a conceptual nature, here the authors can highlight the contributions of Vargo and Lusch (2008a) and Vargo and Lusch (2016), Grönroos and Voima (2013) and Hollebeek et al. (2019). The authors can, furthermore, highlight that SDL and the notion of market relations based on the co-creation *modus operandi* have been the main themes that have occupied the top 10 ranks of the most cited papers within the field.

The authors can furthermore confirm the importance of social media as the contemporary phenomenon shaping our intellectual landscape. While the authors technically could agree on an indigenous inception of social media within the context of the marketing discipline, however, social media (Sheth 2021) amplified by the rise of AI (Davenport et al. 2020) are themes that marketing science still needs to absorb in a way that the authors stay true to Alderson's (1957) call of a marketing definition of utility and value, and in this case of the whole spectrum of social media marketing amplified by the AI technology for marketing. The rise of social media and the notion of digitality is not a new field of or marketing, but moreover, sound marketing practice via the digital channel of social media amplification.

Limitations

The spectrum of contributions of this paper is limited to the two journals of AMS. Additional relevant journals such as the Journal of Marketing (JM) were not a part of this study, while the authors have paved the way for including these journals in possible future research pursuits. The journals relating to consumer behavior (CB) and consumer research based on the Association for Consumer Research (ACR) such as the Journal of Consumer Research (JCR) were also not included. Here the authors construct on Wilkie and Moore (2003), Hunt (2020), and Sheth (1982), as the field of CB has divorced itself from marketing and the foundational premise of this research is to lay a possible future direction for marketing to develop indigenous theory-building capacity. That the first 900 issues of the JCR, mention marketing in the title only 3 times is a clear indication that another rich and vibrant field had evolved, originally within the domain of marketing but now a long-established domain outside the marketing spectrum. Our historical account is founded on the shoulders of the giants of the marketing field and the authors have tried to establish our analysis only on observable and variable themes deriving from the historical accounts, which came before us from a conceptual lens and based on the large data analyzed in this research, which embraces the complete content of JAMS and AMSR (until the issues Volume 49, Issue number 2 of JAMS; Volume 10, Issue number 3 of AMSR). Further, that only author frequencies and not author relationships and co-authorships are reported, could be a limitation.

Discussion

While our research confirms Yadav's (2010; 2020) observations on the nature of the decline of the conceptual articles and the lack of indigenous themes in marketing (Hunt 2018), the authors still have given the evidence that the contributions of conceptual only articles have been essential and vital for the field. In terms of the field's homeostasis and autopoiesis of reviving the doctoral programs with indigenous theories pertaining to marketing science, the authors can deliver hope but at the same time the few "conceptual-only" articles, which could measure themselves on an equal-footing with the quantitative models. Here the authors see also a clear necessity of additional theoretical contributions that could deliver solid impulses to the quantitative-based and consumer-behavior-driven research to expand their themes in the image of foundational marketing aspirations such as the dimensions of establishing sustainability (Lunde 2018) and mindful consumption (Sheth et al. 2011), coping and framing an AI-value definition of marketing and defining solid organizational theories for managing global enterprises, so that effective marketing responses could be dealt with and furthermore to overcome the theoretical limitations of the marketing field (Kamran et al. 2020; Kamran 2021; Key et al. 2020; Sheth 2021; Hunt 2020).

The notion that over 75% of marketing's doctoral students do not self-identify as substantively marketing scholars (Hunt 2020) is a dire warning for all of us in the field (Yadav 2020), especially when the many new doctoral researchers in marketing are not trained to 1) self-identify with the substantive content that may bring the field further; 2) to also not really understand the notion of holistic marketing (Sheth 2021) and to not be trained on the essential methods that would make the field's reproduction of scholars ever possible, and 3) to make the dimension of publish or perish (Wilkie/Moore 2003) as the *raison d'être* of scholarship within the field while not looking at the themes that are promoted so that a majoring in minors is further the state of the art within a very narrow and micro concentration within our field. Marketing is facing a large volume of internal and external troubles (Key et al. 2020; Webster/Lusch 2013; Yadav 2010), thus, the authors can also confirm the deinstitutionalization process of marketing, which is further promoted (Hunt 2020).

Our research confirms that the fields of SCM and CB are not really represented within the contemporary publications and research streams of marketing today. The authors also deliver evidence for exogenous themes and research directions that have shaped the intellectual landscape of marketing in recent years. These streams predominantly include food products, products development, innovation management, CSR, and strategy. The authors can also account for new streams of indigenous research dimensions, especially the contributions of SDL that have shaped the intellectual structure of the field. A huge rise of SDL can be observed between the years 2008-2011. The authors see a constant interest in this field, however, the extreme peak that was observed within the mentioned years has also significantly declined. Additional indigenous research themes are the vital importance of customer experience and the notion of branding that still occupies a solid topology within our interpreted data (see Figure 5).

The authors need furthermore to ask of how marketing could offer a foundation for alignment with the dimensions of shared value (Porter/Kramer 2011) that has brought the whole notion of CSR, which has also been a theme observed to be essential for the field as aforementioned. Furthermore, the global movement and sensitivity of consumers for sustainable products and services and sustainable pricing for offers to the bottom of the pyramid force a much richer ground for innovative approaches towards embracing sustainability.

Here the authors ask that how could the authors bring sustainability broader into the main theme for marketing science in a manner that it fits within the dimension of civilizing capitalism while embracing sustainability as a major research stream in a way that the authors free its spirit from the walls of corporate do-gooding and CSR, towards making being-sustainable truly profitable on a business model proposed by marketing.

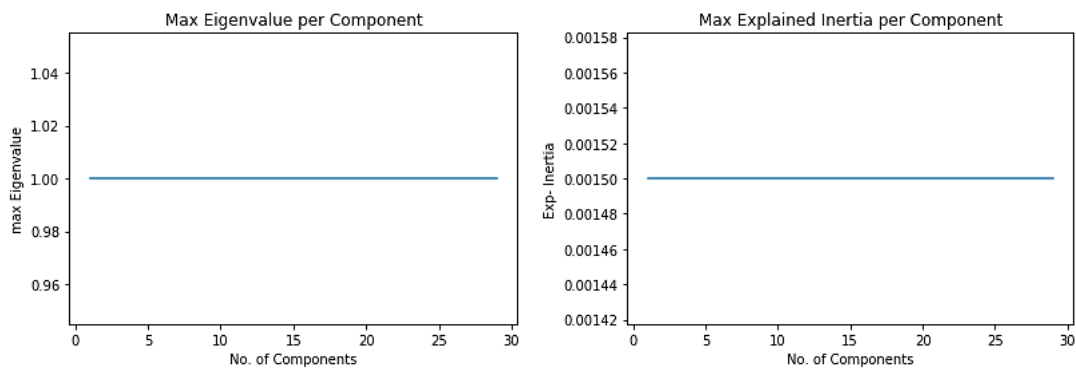
In addition, the authors need to overcome our disjoint with the dimension of organizational theory. While the authors agree with Hult (2011) on the notion of a boundary-spanning marketing organization, the authors want to emphasize the importance of a solid organizational model derived from marketing science, especially in times of rapid disruptions, where short-term but adequate responses are required and agile, models of new-work (Bergmann 2019) remote work, especially in post-Covid 19 society, needs to be designed.

The authors suggest that marketing needs to give directions to the recent developments within the taken place rapprochement in practice between high-tech and marketing, especially observed under the vocation of the “marketing-technologist” (Iansiti/Lakhani 2014; Brinker McLellan 2014). Finally, the authors would like to see marketing acknowledging the contemporary limitation within the marketing theory in terms of the field-defining path-breaking and innovative research streams for marketing so that the field can be aligned and can overcome the state of contemporary unsettledness.

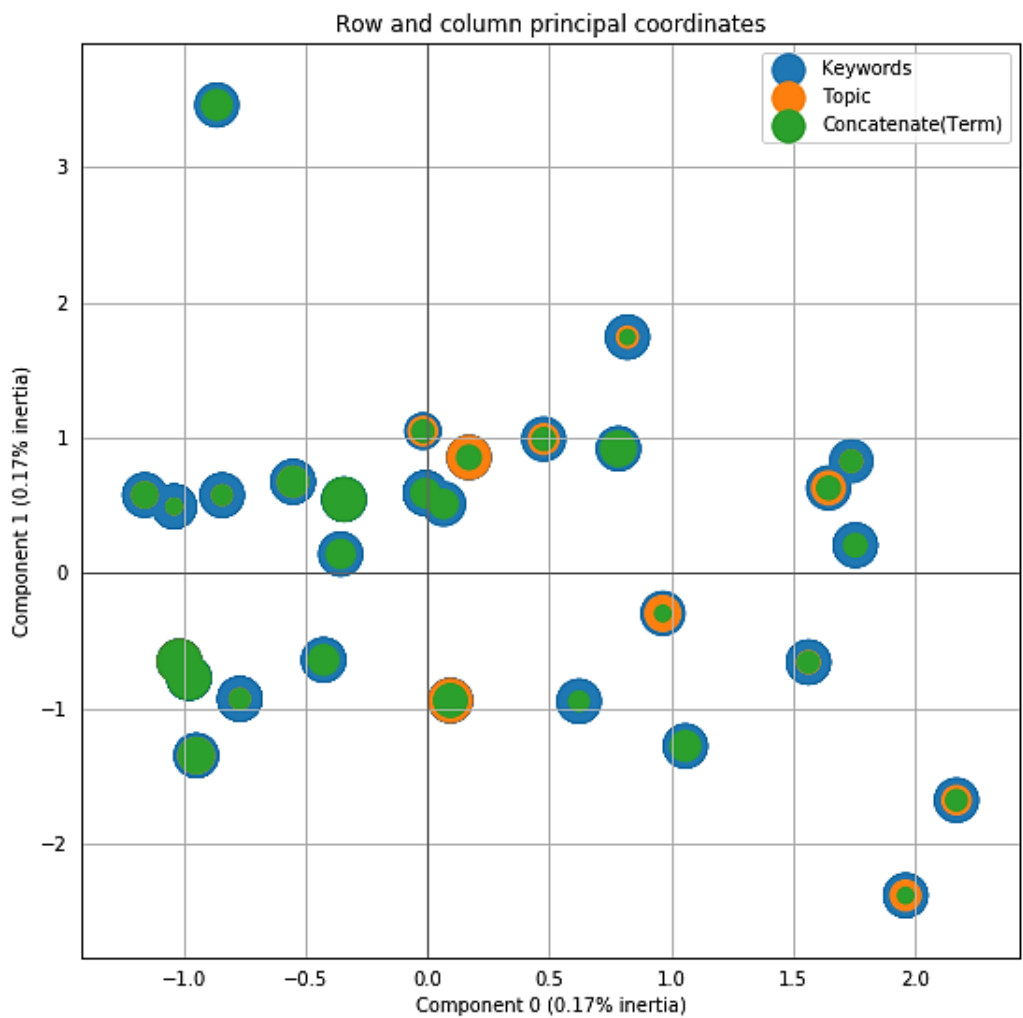
Based on our observations within the research and our experience as marketing community the authors would like to propose another possible solution to Hunt’s (2020) call, thus, instead of focusing on a re-institutionalizational framework for marketing, that marketing science does not only look back at what has been lost but simultaneously for what is to come.

Appendix

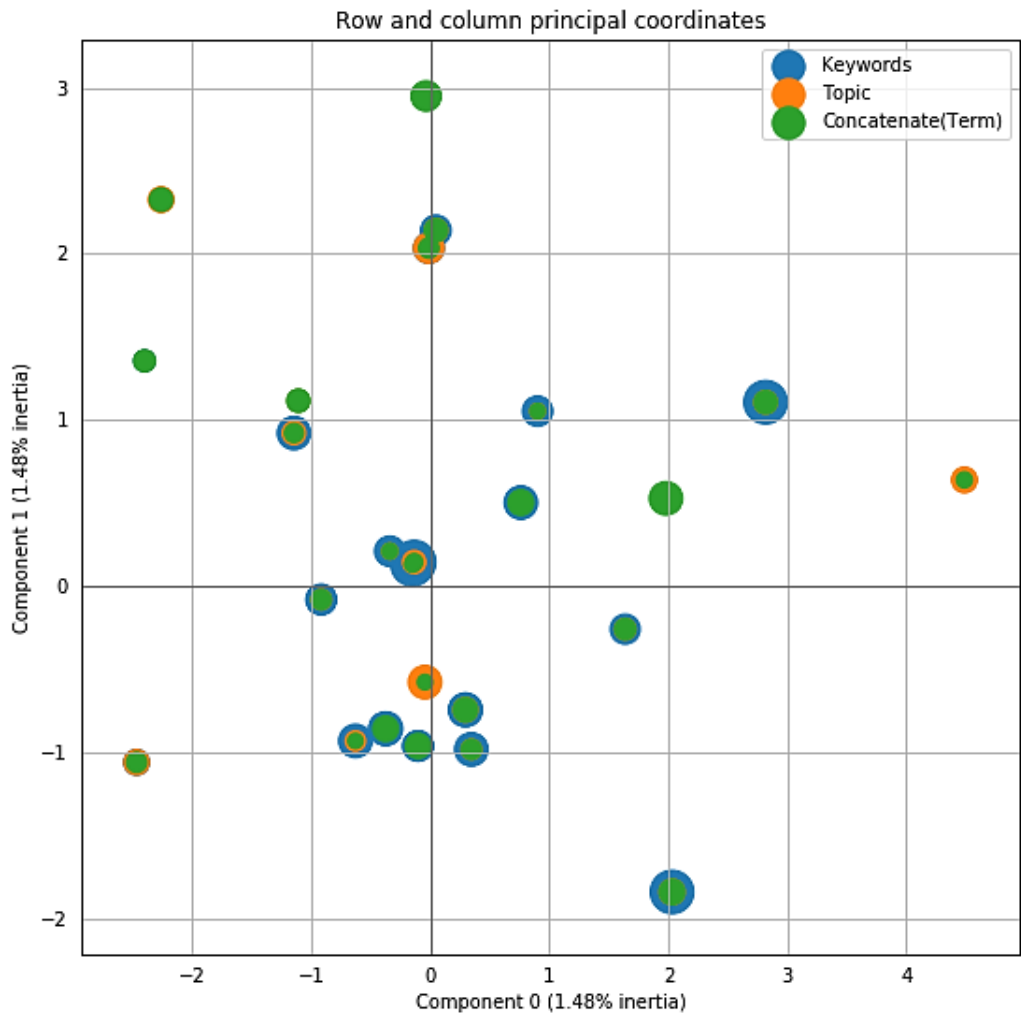
Appendix 1: Sensitivity Analysis of MCA with components $n = 2-30$



Appendix 2: MCA of JAMS



Appendix 3: MCA of AMSR



Appendix 4: Manual Stop List

Marketing, market, markets, research, na, association, business, review, word, life, role, analysis, markets, time, conjoint, study, wall, level, linear, mean, view, study, hedonic, Harvard, journal, studies, customers, consumers, firms, models, results, search, measures, prices, brands, services, countries, paper, relationships, relational, effects, products, cultural, differences, found, significant, ethical, journals, articles, article, intent, approach, field, findings, impact, degree, framework, literature, responses, organizations, drucker, peter, systems, methods, decision, negative, positive, alliances, innovations, students, authors, attributes, programs,

Appendix 5: Association of Main Keywords to Topics

Keywords	Topic
performance, theory, conflict	Performance
channel, structural, distribution	Channel Management
consumer, experience, consumption	Consumer Experience
consumer, behavior, decision	Consumer Behavior
product, data, development	Product Development
food, product, effects	Food Products
shopping, behavior, consumer	Shopping Behavior
performance, firm, orientation	Firm Performance
sales, selling, performance	Sales Management
price, consumer, information	Pricing
industrial, behavior, buying	Industry
advertising, television, commission	Advertising
product, consumer, response	Consumer Response
information, consumer, risk	Consumer Information
consumer, scale, development	Scaling
value, logic, service-dominant	Service Dominant Logic
strategy, theory, strategic	Strategy
international, strategy, foreign	International Marketing
effect, power, model	Model Effects
store, retail, image	Retail
decision, health, healthcare	Healthcare
innovation, management, knowledge	Innovation and Knowledge Management
online, social, media	Social Media
brand, equity, extension	Branding
service, recovery, satisfaction	Customer Satisfaction
response, mail, survey	Surveys
measurement, validity, model	Measurement and Modelling
customer, relationship, service	Customer Relationship
theory, perspective, public	Public Theory
corporate, social, responsibility	Corporate Social Responsibility (CSR)

Appendix 6: KNIME Built-In Stop List

A, about, above, across, after, again, against, all, almost, alone, along, already, also, although, always, among, an, and, another, any, anybody, anyone, anything, anywhere, are, area, areas, around, as, ask, asked, asking, asks, at, away, b, back, backed, backing, backs, be, became, because, become, becomes, been, before, began, behind, being, beings, best, better, between, big, both, but, by, c, came, can, cannot, case, cases, certain, certainly, clear, clearly, come, could, d, did, differ, different, differently, do, does, done, down, downed, downing, downs, during, e, each, early, either, end, ended, ending, ends, enough, even, evenly, ever, every, everybody, everyone, everything, everywhere, f, face, faces, fact, facts, far, felt, few, find, finds, first, for, four, from, full, fully, further, furthered, furthering, furthers, g, gave, general, generally, get, gets, give, given, gives, go, going, good, goods, got, great, greater, greatest, group, grouped, grouping, groups, h, had, has, have, having, he, her, here, herself, high, high, high, higher, highest, him, himself, his, how, however, i, if, important, in, interest, interested, interesting, interests, into, is, it, its, itself, j, just, k, keep, keeps, kind, knew, know, known, knows, l, large, largely, last, later, latest, least, less, let, lets, like, likely, long, longer, longest, m, made, make, making, man, many, may, me, member, members, men, might, more, most, mostly, mr, mrs, much, must, my, myself, n, necessary, need, needed, needing, needs, never, new, newer, newest, next, no, nobody, non, noone, not, nothing, now, nowhere, number, numbers, o, of, off, often, old, older, oldest, on, once, one, only, open, opened, opening, opens, or, order, ordered, ordering, orders, other, others, our, out, over, p, part, parted, parting, parts, per, perhaps, place, places, point, pointed, pointing, points, possible, present, presented, presenting, presents, problem, problems, put, puts, q, quite, r, rather, really, right, room, rooms, s, said, same, saw, say, says, second, seconds, see, seem, seemed, seeming, seems, sees, several, shall, she, should, show, showed, showing, shows, side, sides, since, small, smaller, smallest, so, some, somebody, someone, something, somewhere, state, states, still, such, sure, t, take, taken, than, that, the, their, them, then, there, therefore, these, they, thing, things, think, thinks, this, those, though, thought, thoughts, three, through, thus, to, today, together, too, took, toward, turn, turned, turning, turns, two, u, under, until, up, upon, us, use, used, uses, v, very, w, want, wanted, wanting, wants, was, way, ways, we, well, wells, went, were, what, when, where, whether, which, while, who, whole, whose, why, will, with, within, without, work, worked, working, works, would, x, y, year, years, yet, you, young, younger, youngest, your, yours, z,

References

- Alderson, W. (1957): *Marketing Behavior and Executive Action*, Richard D. Irwin. Inc.; Homewood, IL.
- Alderson, W.; Wooliscroft, B.; Tamilia, R. D. & Shapiro, S. J. (2006): *A twenty-first century guide to Aldersonian marketing thought*. New York, NY: Springer Science+Business Media.
- Ashby, W. R. (1956): *An Introduction to Cybernetics. Part Two: Variety*. London: Methuen.
- Attias, H.; Platt, J.; Acero, A. & Deng, L. (2000): Speech denoising and dereverberation using probabilistic models. *Advances in neural information processing systems*, 13.
- Bagozzi, R. P. (1994): *Advanced Methods of Marketing Research*. Cambridge, Mass.: Blackwell.
- Barnard, N. D. (2019): Ignorance of Nutrition Is No Longer Defensible. *JAMA internal medicine*, 179, 1021–1022. doi:10.1001/jamainternmed.2019.2273.
- Bartels, R. (1962): *The development of marketing thought*. Homewood, Illinois: Irwin.
- Bartels, R. (1967): A Model for Ethics in Marketing. *Journal of Marketing*, 31, 20–26. doi:10.1177/002224296703100105.
- Bartels, R. (1974): The Identity Crisis in Marketing. *Journal of Marketing*, 38, 73–76. doi:10.1177/002224297403800413.
- Bartels, R. (1988): *The history of marketing thought (3rd ed.)*. Columbus, Ohio: Publishing Horizons.
- Beer, S. (1959): *Cybernetics and management (Management science series)*. London: English University Press.
- Beer, S. (1971): The liberty machine. *Futures*, 3, 338–348. doi:10.1016/0016-3287(71)90053-X.
- Benzecri, J.-P. (1982): *L'analyse des donnees, tome 2: L'analyse des correspondances*.
- Bergh, D. D.; Perry, J. & Hanke, R. (2006). Some predictors of SMJ article impact. *Strategic Management Journal*, 27, 81–100. doi:10.1002/smj.504.
- Bergmann, F. (2019): *New work, new culture: Work the authors want and a culture that strengthens us* / Frithjof Bergmann. Winchester, UK: Zero Books.
- Berry, L. L. & Parasuraman, A. (1993): Building a new academic field – The case of services marketing. *Journal of Retailing*, 69, 13–60. doi:10.1016/S0022-4359(05)80003-X.
- Bertalanffy, L. von. (1969): *General system theory: Foundations, development, applications*.
- Blei, D. M. (2012): Probabilistic topic models. *Communications of the ACM*, 55(4), 77–84.
- Blei, D. M.; Ng, A. Y. & Jordan, M. I. (2003): Latent dirichlet allocation. *the Journal of machine Learning research*, 3, 993–1022.
- Boorstin, D. J. (1978): *Republic of technology: reflections on our future community*. New York, NY: Harper and Row.
- Borden, N. H. (1964): The concept of the marketing mix. *Journal of Advertising Research*, 4(2), 2–7.
- Brinker, S. & McLellan, L. (2014): The rise of the chief marketing technologist. *Harvard Business Review*, 92(7), 82–85.

- Davenport, T.; Guha, A.; Grewal, D. & Bressgott, T. (2020): How artificial intelligence will change the future of marketing. *Journal of the Academy of Marketing Science*, 48, 24–42. doi:10.1007/s11747-019-00696-0.
- Dickey, J. M. (1983): Multiple hypergeometric functions: Probabilistic interpretations and statistical uses. *Journal of the American Statistical Association*, 78(383), 628–637.
- El-Ansary, A.; Shaw, E. H. & Lazer, W. (2018): Marketing's identity crisis: insights from the history of marketing thought. *AMS Review*, 8, 5–17. doi:10.1007/s13162-017-0102-y.
- Ellen, P. S. (2006): Building Corporate Associations: Consumer Attributions for Corporate Socially Responsible Programs. *Journal of the Academy of Marketing Science*, 34, 147–157. doi:10.1177/0092070305284976.
- Furrer, O.; Thomas, H. & Goussevskaia, A. (2008): The structure and evolution of the strategic management field: A content analysis of 26 years of strategic management research. *International Journal of Management Reviews*, 10, 1–23. doi:10.1111/j.1468-2370.2007.00217.x.
- Gardner, B. B. & Levy, S. J. (1955): The product and the brand. *Harvard Business Review*, 33(2), 33–39.
- Grether, E. T. (1976): The First Forty Years. *Journal of Marketing*, 40, 63–69. doi:10.1177/002224297604000315.
- Grönroos, C. (1982): An Applied Service Marketing Theory. *European Journal of Marketing*, 16, 30–41. doi:10.1108/EUM0000000004859.
- Grönroos, C. (2006): Adopting a service logic for marketing. *Marketing Theory*, 6, 317–333. doi:10.1177/1470593106066794.
- Grönroos, C. & Voima, P. (2013): Critical service logic: making sense of value creation and co-creation. *Journal of the Academy of Marketing Science*, 41, 133–150. doi:10.1007/s11747-012-0308-3.
- Gummesson, E. (1979): The Marketing of Professional Services – An Organizational Dilemma. *European Journal of Marketing*, 13, 308–318. doi:10.1108/EUM0000000004951.
- Gummesson, E. (2008): Extending the service-dominant logic: from customer centricity to balanced centricity. *Journal of the Academy of Marketing Science*, 36, 15–17. doi:10.1007/s11747-007-0065-x.
- Heck, J. L. & Cooley, P. L. (1988): Most Frequent Contributors to the Finance Literature. *Financial Management*, 17, 100. doi:10.2307/3666076.
- Hoffman, D. L. & Franke, G. R. (1986): Correspondence Analysis: Graphical Representation of Categorical Data in Marketing Research. *Journal of Marketing Research*, 23, 213–227. doi:10.1177/002224378602300302.
- Hollebeek, L. D.; Srivastava, R. K. & Chen, T. (2019): S-D logic-informed customer engagement: integrative framework, revised fundamental propositions, and application to CRM. *Journal of the Academy of Marketing Science*, 47, 161–185. doi:10.1007/s11747-016-0494-5.
- Hult, G. T. M. (2011): Toward a theory of the boundary-spanning marketing organization and insights from 31 organization theories. *Journal of the Academy of Marketing Science*, 39, 509–536. doi:10.1007/s11747-011-0253-6.

- Hult, G. T. M. (2015): JAMS 2010–2015: literature themes and intellectual structure. *Journal of the Academy of Marketing Science*, 43, 663–669. doi:10.1007/s11747-015-0464-3.
- Hunt, S. D. (2018): Advancing marketing strategy in the marketing discipline and beyond: from promise, to neglect, to prominence, to fragment (to promise?). *Journal of Marketing Management*, 34(1-2), 16–51.
- Hunt, S. D. (2020): For re-institutionalizing the marketing discipline in Era V. *AMS Review*, 10, 189–198. doi:10.1007/s13162-020-00183-8.
- Iansiti, M. & Lakhani, K. R. (2014): Digital ubiquity: How connections, sensors, and data are revolutionizing business. *Harvard Business Review*, 92(11), 19.
- Jones, D. G. B. & Monieson, D. D. (1990): Historical research in marketing: Retrospect and prospect. *Journal of the Academy of Marketing Science*, 18, 269–278. doi:10.1007/BF02723911.
- Kaciak, E. & Louviere, J. (1990): Multiple Correspondence Analysis of Multiple Choice Experiment Data. *Journal of Marketing Research*, 27, 455. doi:10.2307/3172630.
- Kamran, Q. (2021): Managing complexity in marketing: From a design Weltanschauung. Enschede: University of Twente.
- Kamran, Q.; van Dijk, J.; Topp, S. & Henseler, J. (2020): The Evolving New Topology of Marketing From A Design Weltanschauung. *AMA Summer Academic Conference Bridging Gaps: Marketing in an Age of Disruption August 18-20, 2020*, 997–1012.
- Key, T. M.; Clark, T.; Ferrell, O. C.; Stewart, D. W. & Pitt, L. (2020): Marketing's theoretical and conceptual value proposition: opportunities to address marketing's influence. *AMS Review*, 10, 151–167. doi:10.1007/s13162-020-00176-7.
- Kotler, P. (1967): *Marketing Management: Analysis, Planning, and Control*. Prentice-Hall.
- Kotler, P. (1972): A Generic Concept of Marketing. *Journal of Marketing*, 36, 46–54. doi:10.1177/002224297203600209.
- Kotler, P. (2005): The Role Played by the Broadening of Marketing Movement in the History of Marketing Thought. *Journal of Public Policy & Marketing*, 24, 114–116. doi:10.1509/jppm.24.1.114.63903.
- Kotler, P. & Levy, S. J. (1969): Broadening the Concept of Marketing. *Journal of Marketing*, 33, 10–15. doi:10.1177/002224296903300103.
- Lebart, L. (1984). *Multivariate Descriptive Statistical Analysis: Correspondence Analysis and Related Techniques for Large Matrices*.
- Levitt, T. (1960). Marketing myopia. *Harvard Business Review*, 38(4), 24–47.
- Levitt, T. (1983): The Globalization of Markets. *Harvard Business Review*. 61(3), 92–101.
- Levy, S. J. & Kotler, P. (1979): Toward a broader concept of marketing's role in social order. *Journal of the Academy of Marketing Science*, 7, 233–238. doi:10.1007/BF02721877.
- Luck, D. J. (1969): Broadening the Concept of Marketing. Too Far. *Journal of Marketing*, 33, 53. doi:10.2307/1248482.

- Lunde, M. B. (2018): Sustainability in marketing: a systematic review unifying 20 years of theoretical and substantive contributions (1997-2016). *AMS Review*, 8(3-4), 85–110.
- Maignan, I. & Ferrell, O. C. (2004): Corporate Social Responsibility and Marketing: An Integrative Framework. *Journal of the Academy of Marketing Science*, 32, 3–19. doi:10.1177/0092070303258971.
- McCallum, A. (2002): MALLT: A Machine Learning for Language Toolkit. <http://mallet.cs.umass.edu/>. Accessed 21 May 2021.
- McCarthy, E. J. (1960): *Basic marketing: A managerial approach*. Homewood: Richard D. Irwin, Inc.
- McKitterick, J. (1957): What is the marketing management concept? In: Bass, F. M. (ed.), *The frontiers of marketing thought and science*. Chicago: American Marketing Association.
- Newman, D.; Asuncion, A.; Smyth, P. & Welling, M. (2009): Distributed algorithms for topic models. *Journal of Machine Learning Research*, 10(8).
- Porter, M. E. & Kramer, M. R. (2011): The Big Idea: Creating Shared Value. How to reinvent capitalism – and unleash a wave of innovation and growth. *Harvard Business Review*, 89(1–2).
- Powers, T. L. (2012): Early schools of marketing thought and marketplace evolution. *Journal of Historical Research in Marketing*, 4, 190-206. doi:10.1108/17557501211195127.
- Ramos-Rodríguez, A.-R. & Ruíz-Navarro, J. (2004): Changes in the intellectual structure of strategic management research: a bibliometric study of the *Strategic Management Journal*, 1980–2000. *Strategic Management Journal*, 25, 981–1004. doi:10.1002/smj.397.
- Sen, S. (2006): The Role of Corporate Social Responsibility in Strengthening Multiple Stakeholder Relationships: A Field Experiment. *Journal of the Academy of Marketing Science*, 34, 158–166. doi:10.1177/0092070305284978.
- Shaw, E. H. (1995): The First Dialogue on Macromarketing. *Journal of Macromarketing*, 15, 7–20. doi:10.1177/027614679501500103.
- Shaw, E. H. (2009): Reflections on The History of Marketing Thought. *Journal of Historical Research in Marketing*, 1, 330-345. doi:10.1108/17557500910974640.
- Shaw, E. H. & Jones, D. G. (2005): A history of schools of marketing thought. *Marketing Theory*, 5, 239–281. doi:10.1177/1470593105054898.
- Sheth, J. (2021). New areas of research in marketing strategy, consumer behavior, and marketing analytics: the future is bright. *Journal of Marketing Theory and Practice*, 29, 3–12. doi:10.1080/10696679.2020.1860679.
- Sheth, J. N. (1982): *Consumer Behavior: Surpluses & Shortages*. ACR North American Advances.
- Sheth, J. N.; Gardner, D. M. & Garret, D. E. (1988): *Marketing theory : evolution and evaluation*. New York: John Wiley.
- Sheth, J. N.; Sethia, N. K. & Srinivas, S. (2011): Mindful consumption: a customer-centric approach to sustainability. *Journal of the Academy of Marketing Science*, 39, 21–39. doi:10.1007/s11747-010-0216-3.
- Shostack, G. L. (1977): Breaking Free from Product Marketing. *Journal of Marketing*, 41, 73–80. doi:10.1177/002224297704100219.

- Skålén, P.; Gummerus, J.; von Koskull, C. & Magnusson, P. R. (2015): Exploring value propositions and service innovation: a service-dominant logic study. *Journal of the Academy of Marketing Science*, 43, 137–158. doi:10.1007/s11747-013-0365-2.
- Smith, W. R. (1956): Product Differentiation and Market Segmentation as Alternative Marketing Strategies. *Journal of Marketing*, 21, 3–8. doi:10.1177/002224295602100102.
- Tahai, A. & Meyer, M. J. (1999): A revealed preference study of management journals' direct influences. *Strategic Management Journal*, 20, 279–296. doi:10.1002/(SICI)1097-0266(199903)20:3<279:AID-SMJ33>3.0.CO;2-2.
- Tamilia, R. D. (2011): Reflections on the history of marketing thought and theory development. *Marketing Theory*, 11, 507–512. doi:10.1177/1470593111418808.
- Teh, Y. W.; Jordan, M. I.; Beal, M. J. & Blei, D. M. (2006): Hierarchical dirichlet processes. *Journal of the American Statistical Association*, 101(476), 1566–1581.
- TutorialsPoint. (2021): OpenNLP – Tokenization. https://www.tutorialspoint.com/opennlp/opennlp_tokenization.htm. Accessed 21 May 2021.
- Vargo, S. L. & Lusch, R. F. (2004): Evolving to a New Dominant Logic for Marketing. *Journal of Marketing*, 68, 1–17. doi:10.1509/jmkg.68.1.1.24036.
- Vargo, S. L. & Lusch, R. F. (2008a): Service-dominant logic: continuing the evolution. *Journal of the Academy of Marketing Science*, 36, 1–10. doi:10.1007/s11747-007-0069-6.
- Vargo, S. L. & Lusch, R. F. (2008b): Why “service”? *Journal of the Academy of Marketing Science*, 36, 25–38. doi:10.1007/s11747-007-0068-7.
- Vargo, S. L. & Lusch, R. F. (2016): Institutions and axioms: an extension and update of service-dominant logic. *Journal of the Academy of Marketing Science*, 44, 5–23. doi:10.1007/s11747-015-0456-3.
- Vargo, S. L. & Lusch, R. F. (Eds.). (2006): *The Service-Dominant Logic of Marketing*. Armonk, NY: M.E. Sharpe Inc.
- Vargo, S. L. & Morgan, F. W. (2005): Services in Society and Academic Thought: An Historical Analysis. *Journal of Macromarketing*, 25, 42–53. doi:10.1177/0276146705275294.
- Vargo, S. L.; Akaka, M. A. & Vaughan, C. M. (2017): Conceptualizing value: a service-ecosystem view. *Journal of Creating Value*, 3(2), 117–124.
- Vargo, S. L.; Lusch, R. F. & Morgan, F. W. (2006): Historical perspectives on service-dominant logic. In S. L. Vargo & R. F. Lusch (Eds.), *The Service-Dominant Logic of Marketing* (pp. 29–42). Armonk, NY: M.E. Sharpe Inc.
- Webster, F. E. & Lusch, R. F. (2013): Elevating marketing: Marketing is dead! Long live marketing! *Journal of the Academy of Marketing Science*, 41.
- Wiener, N. (1948): *Cybernetics: Or control and communication in the animal and the machine* (The technology press). New York, NY: Wiley.
- Wilkie, W. L. & Moore, E. S. (2003): Scholarly Research in Marketing: Exploring the “4 Eras” of Thought Development. *Journal of Public Policy & Marketing*, 22, 116–146. doi:10.1509/jppm.22.2.116.17639.

- Yadav, M. S. (2010). The Decline of Conceptual Articles and Implications for Knowledge Development. *Journal of Marketing*, 74(1), 1–19. <https://doi.org/10.1509/jmkg.74.1.1>.
- Yadav, M. S., (2020), Reimagining marketing doctoral programs, *AMS Review*, 10, issue 1, p. 56–64, https://EconPapers.repec.org/RePEc:spr:amsrev:v:10:y:2020:i:1:d:10.1007_s13162-020-00169-6
- Yao, L.; Mimno, D. & McCallum, A. (2009): Efficient methods for topic model inference on streaming document collections. In New York, New York, USA. New York, NY: ACM. doi:10.1145/1557019.1557121.

Authors

Prof. Dr. Dr. Qeis **Kamran's** career, spanning three decades, is distinguished by top roles in energy, consulting, and education. He holds multiple Ph.D.s in Marketing, AI, Design Science, Strategy Management, Cybernetics, and Systems Theory, and MBAs in International General Management, Projects, and Process Management. With law degrees from St. Gallen and Augsburg, Prof. Kamran has excelled as a professor, researcher, senior lecturer, and leader in consulting, training, and entrepreneurship, also contributing significantly to international journals.



Prof. Dr. Marcus **Becker** is a full-time professor of “Quantitative Methods of Business Information Technology” at the International School of Management (ISM) in Dortmund. He gained international experience in automated Risk Management processes working for HSBC and Deloitte. Besides, he was part of the German Robo Advisory panel at Deloitte’s Risk Advisory service line testing Machine Learning algorithms like LSTM on stock market forecasts.



Becker, Marcus; Prokop Dayrell de Lima, Erika

Natural Language Processing in IT Ticketing Systems

A conceptual framework for Question-and-Answering machines based on GPT-Algorithms

Abstract

This research is a feasibility study to design a Natural Language Processing (NLP) system within a Question-and-Answering (Q&A) environment for internal IT help desk ticketing operations. The proceedings will be used to develop a conversational agent for an IT consultancy company. Tests with few-shot learning algorithms were performed by calibrating two GPT-2 language models. Another benchmark model was tested on the most recent GPT-3 standard. The input data stems from software license release requests of an internal IT help desk. The final model will be a hybrid approach of first guidance by an automated agent and a human expert intervention for more complicated IT problem. The agent will improve itself by constantly evaluating user feedback.

Keywords:

Natural Language Processing, Question Answering, GPT-2, GPT-3, Few-shot learning, Natural Language Generation, Conversational Agent Implementation

JEL Classification Numbers: C67, C88, M15, 031, 032

Acknowledgments

We would like to thank all the experts who contributed to our research with their valuable feedback. Special thanks to Jörg Wiethölter, Jürgen Angele, and Sascha Schramm.

1 Introduction

The human-machine interaction is gaining more and more awareness in daily business operations – especially in times of ongoing digitalization processes. This interaction is not without flaws. In case of disruptions, the IT help desk usually is the first point of contact for internal employees who are experiencing IT problems. Interruptions of work processes are costly for the company. Apart from this, they pose a potential source of accumulating frustration on user side. Therefore, it is pivotal to react promptly to user requests. According to a research conducted by “Forrester Research” in 2020 organizations that have adopted help desk services, increased their overall employee satisfaction and managed to achieve on average a self-service resolution rate of 20%.¹ IBM, the global market leader in IT infrastructure solutions, sees the following major contributions of IT help desks: Service desks standardize IT processes, and automate routine tasks, thus increasing operational efficiency. They align users with IT processes by offering self-service tools and automated workflows that further increase business productivity and free IT staff to focus on more strategic activities. Service desks act as a “hub” for all hardware and software assets in the company. Moreover, they create digitalized contracts, and enable easy information retrieval, IT configuration, and IT maintenance. Finally, IT service desks keep business activities up in times of technical riots, thus avoiding costly downtimes in operations.²

Due to increasing demand for technical support, the IT help desk has to organize its operations in form of a so-called “ticketing system”. A majority of tickets, like software licensing issues, are constantly recurring problems, with pre-existing solutions. Hence, it would constitute a tremendous relief for IT specialists, if such easy-to-solve tickets could be resolved automatically, e.g. in form of a conversational agent, thereby releasing the costlier human IT resources.

The worldwide adoption of the IT Infrastructure Library framework (ITIL) has already shown some effectiveness in improving service desk operations. However, successive improvements, for many IT service providers, still remain a major challenge (Jäntti et al. 2012). The goal of IT service desk agents is to record, classify, diagnose, and resolve user requests (Jäntti et al. 2012). Requests consist of the following: software and hardware failures, requests for resetting passwords, requests for new software packages and licenses, or change requests for existing services. Finally, the help desk constantly receives feedback for its operations. Feedback can be positive as well as negative. An important step in improving operations would be to ensure effective communication between the service desk and the user. This includes a continuous overview of the progress of ticket resolution, which has to be preserved through all possible channels (Jäntti et al. 2012). Moreover, timely ticket resolution increases user satisfaction.

¹ See Charles Betz, Andrew Hewitt, “Service Desk 2020: It’s All About Employee Experience,” Forrester Research, October 6, 2020.

² See <https://www.ibm.com/topics/service-desk>, last accessed on 12.07.2023, 16: 46.

In this study, we design a feasibility study for a Natural Language Processing (NLP) system which operates within a Question-and-Answer (Q&A) environment for first-level IT support. By using the most recently developed Generative Pre-trained Transformer technology (GPT) as a few-shot learning method. Few-shot learning scenarios originate from past IT ticket resolutions. For calibration purposes, we use the pre-implemented open-source-based, *GPT-2 packages gpt-2-simple and german-gpt2*. In a later run of our analyses, we were also granted access to the most recently developed GPT-3 standard called *Playground*.

Task-oriented dialog systems have encountered increased awareness since they can assist users in various, predominantly monotonous, tasks. In a “typical” task-oriented dialog system, the Natural Language Generation (NLG) module converts a system action (e.g., often specified in a semantic form selected by a dialog policy) into a final response (Peng et al. 2020). The response should be adequate, fluent and represent a semantic dialog action to engage users’ attention. Existing methods for NLG can be subdivided into two major categories. (1) Template-based methods require domain expert knowledge to create templates for each domain, with the system filling in so-called “slot values” (Cheyer and Guzzoni, 2014; Langkilde and Knight, 1998). (2) Statistical language models, as presented by Gao et al. (2019), can be built on neural networks. The models are trained on a labeled text corpus. One canonical model is semantically conditioned, such as SC-LSTM (Wen et al. 2015), which encodes dialog acts with one-hot representations and uses it as an extra feature to inform the sentence generation process. The general performance is good but the training amount needs a large amount of domain-specific data. Moreover, it suffers from scalability as dialog combinations grow exponentially.

During our study, we collaborated with a large-scale IT consultancy positioned in Dortmund (North Rhine-Westphalia, Germany). The company already has a pre-existing conversational agent solution that could be enriched by our GPT module. For this, we need a structured example file of previously resolved tickets. Our first tests were appealing with respect to human (qualitative) judgment. The answers fitted well to the respective questions. Even first guidance advices were provided. To give an example: After requesting a license for a specific software product, the agent informed the employee about the costs and the need for approval by a Project Leader or a Manager (see figure 7 in the final evaluation report section). The next step would be to program a conversational agent which directly sends an email to the manager asking for approval. We do not claim that our tool has the ability, nor the intent, to replace skilled human IT resources. But as our anecdotal evidence shows, IT support vastly suffers from work overload. This overload trend will much likely aggravate due to an increasing need for internal digitalization projects and decreasing numbers of skilled workers in the nearby future. Therefore, our work represents a feasibility study that aims to explain how NLP systems can be used to relieve human IT support resources for more complex IT projects. Projects which cannot be resolved by an automated agent.

As prior research conducted by Tarlaci (2019) indicates, unsupervised language models' like GPT-2 have a proud capability to optimize automated Q&A tasks with minimal or no task-specific adaptation. The first public release of GPT-2 (with 124 M parameters) was fine-tuned for a special Q&A task, using the Stanford Question Answering Data-set (SQuAD) 2.0, Huggingface's PyTorch implementation and a modified BERT as Q&A model. The fine-tuned model did not meet the researcher's expectations, regardless of further adjustments on tokenization, hyperparameter fine-tuning and increasing number of epochs. The author indicates that the size of the language model, as well as the complex nature of Q&A tasks in general, are two limiting factors for future deployments of such model. The author further concluded that a more sophisticated fine-tuning experiment, with more performant extraction systems, could improve future results.

Fiore et al. (2020) emphasize the unexploited potential for conversational agents inside a company as Virtual Enterprise Assistants (VEA). Focusing on VEA for in-company IT help desks, a user study was performed with a functional conversational agent prototype to evaluate the suitability for typical IT support cases, e.g. in password resets, answers to Frequently-Asked-Questions (FAQs), and release email attachments from quarantine. In accordance to that, the researchers studied the potential user acceptance. The study was conducted with 12 employees from two companies, a bank and a hospital, that were specifically chosen due to their IT affinity. The results indicate that such an in-company conversational agent is well-perceived by employees. However, the less structured task for answering FAQs aroused negative remarks regarding misclassifications of inputs. Some users had to rephrase their original request since the conversational agent did not understand their first request.

A consecutive work of Fiore et al. (2020) underlines the capabilities of conversational agents to relieve employees of repetitive work in in-house support. Moreover, the authors indicate success conditions for embedding conversational agents in the company's IT landscape. First, it helps to mention that the agent is not human at the beginning of each conversation. Second, a direct offer to escalate the users' concerns to a real human person should be provided. Finally, the authors point out that special Natural Language Understanding (NLU) techniques are necessary for a successful development and maintenance of conversational agents, a shortcoming that most likely has been overcome by the most recently developed ChatGPT 3.5 technology.

A pioneer study of exploiting state-of-the-art pre-trained language models for NLG in task-oriented dialog systems was performed by Peng et al. (2020). The model is semantically-conditioned on GPT-2, named SC-GPT, and built upon Huggingface PyTorch Transformer technology with 345 M parameters. It follows a three-stage training procedure: first, like GPT-2, massive plain language pre-training, second an additional controlled pre-training stage on vast amounts of dialog-act labeled data, in order to allow the generation of texts, third a fine-tuning stage, on a specific task with a limited amount of domain labels. According to Peng et al. (2020) the results indicate that with minimal training examples,

SC-GPT is able to generate adequate natural language responses. The authors trained their model on a new benchmark dataset, called “FEWSHOT-WOZ” based on the “MultiWOZ” dataset developed by Budzianowski et al. (2018) and the “Cambridge NLG datasets” of Wen et al. (2016). The resulting dataset consists of seven domains (Restaurant, Laptop, Hotel, TV, Attraction, Train, Taxi). Although, the domain of “Laptop” might have some references relating to IT help desk topics, we doubt, that the model would be feasible to handle real-life IT help desk problems on a daily basis, especially for highly specialized IT topic for a technical consultancy firm. Moreover, as most scientific language models, the model presented in Peng et al. (2020) is calibrated on stylized data that has been pre-processed with additional speech tags that alleviates the model training but needs additional man power to either manually incorporate speech tags, or in case of automated speech tagging, binds some resources of validating these tags. Our feasibility study aims at developing model that incorporates some “Natural Language Understanding” (NLU) components that do not need additional tagging efforts.

To the best of our knowledge, the realizations of conversational agents in real-life companies using the most recent GPT technology are sparse. On top, IT ticketing services pose an extra challenge due to the complexity of the underlying IT problems, especially, for the higher demands of an IT consultancy.

Our research paper is organized as follows. Section 2 presents a general overview of our model architecture. In Section 3, we develop our main research question, i.e. our objective to implement a GPT-based NLP system for Q&A tasks in IT support processes. Further, we design a feasibility study on how to incorporate this NLP system in a large-scale IT consultancy company. We provide a technical guideline and highlight methodological questions, which should be resolved before starting with the IT implementation. Section 5 outlines the model limitations and opens up for the debate in future research. Section 6 summarizes the paper.

2 Natural Language Processing in Question-Answering-Systems

2.1 General model architecture

A very general conversational agent architecture can be partitioned into the following components: NLU, Dialogue Management and an additional NLG component (Adamopoulou/Moussiades 2020). Taking into consideration the pipeline architecture represented in figure 1, where each Q&A turn is processed by employing a five-step procedure. The transcriptions of user inputs are first processed by the NLU module, where the user intention and other key information are extracted. Afterwards, the information is formatted to the dialog management component, i.e. question, statement, or a request for action, is the input to the NLG, through which a system response in natural language is generated (Peng et al. 2020). Finally, the answer is provided to the user by text, voice or image, according to the interaction methods used in the conversational agent. The agent asks follow-up questions after the

intent is recognized (Kucherbaev et al. 2018). It can be subdivided into a dialogue state tracking component, responsible for monitoring the mapped intent (and entities) until that point of the conversation, and a dialog policy component, which produces a dialog act based on the fact (or entities) retrieved from external resources, such as database or a knowledge base (Peng et al. 2020).

A database serves as part of the backend, where all relevant information is retrieved. Once the appropriate information is extracted. It is forwarded to the Dialog Management Module and then to the Response Generation Module (Adamopoulou/Moussiades 2020). The response to the user is generated based on the dialog act emitted by the Dialog Policy Module. Several approaches can be utilized for generating the user response, including Rule-based, Retrieval-based, and Generative-based methods. In the Rule-based model, the response is selected from a predefined set of rules without generating new text responses. The Retrieval-based model offers more flexibility by selecting the most suitable response through the examination and analysis of available resources (Adamopoulou/Moussiades 2020). The Generative model employs Natural Language Generation (NLG) to generate responses in a human-like natural language manner, considering the previous inputs (Peng et al. 2020; Adamopoulou/Moussiades 2020). Hybrid conversational agents combine a rule-based or retrieval-based approach with a generative approach to handle user inputs that do not match any predefined rules mentioned earlier. Furthermore, the operation of a conversational agent can involve human intervention as well (Mathur/Lopez 2019; Adamopoulou/Moussiades 2020).

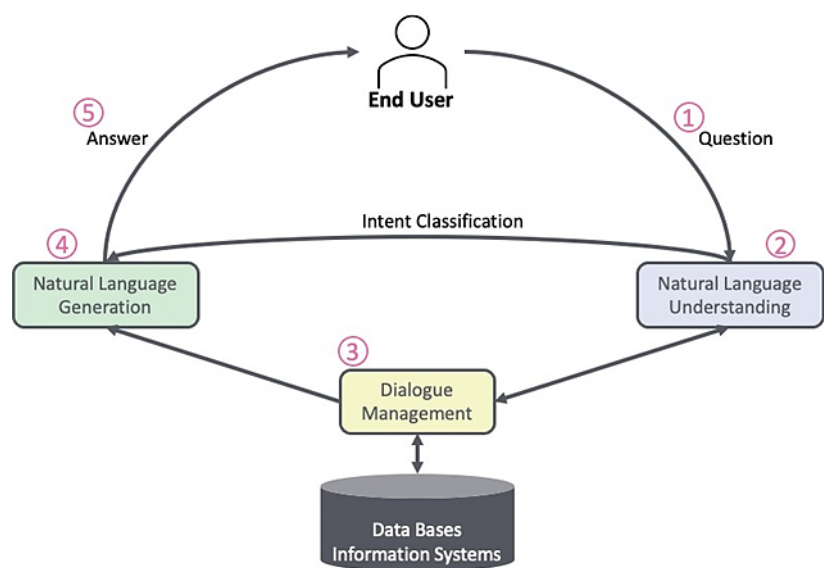


Figure 1: General Architecture of Q&A system based on text generation
Source: Hirschman/Gaizauskas 2001; Peng et al. 2020; Adamopoulou/Moussiades 2020

2.2 GPT Technology

2.2.1 General Framework of GPT

The Generative Pre-trained Transformers (GPT) are a series of language models developed from the Transformer model. Transformer models are a series of connected encoder and decoder nodes (Vaswani et al. 2017). Thus, they belong to a most recently developed sub-category of Deep learning models that are pre-dominantly used in the field of NLP. Transformers thereby replace the previously used Recurrent Neural Network (RNN) models, such as Long-Short-Term-Memory (LSTM), and Convolutional Neural Networks (CNN) models.

GPT is an NLP model based on unsupervised machine learning methods. Without any particular task-specific training and on any data subject, the GPT-2 model can read and “understand” a text, summarize or transcribe it. GPT-3 is able to answer questions about the information contained in it. Moreover, it is capable of completing, or even generating paragraphs with grammatical and syntactic consistency. GPT- assigns a probability to a sequence of words for a given size. It obtains a probability of occurrence for each sequence determined by previously seen training data. Training data is collected and retrieved from an extensive number of web sources. The notable innovation of GPT-2 consists in its learning method: “*Zero-shot-learning*” covers a technology where text, that has never been learned before, is inserted, thus starting from zero, and which evaluates answer’s consistency. Instead of rearranging the sequences for fine-tuning, as previously conducted by GPT-1, the input to GPT-2 provides a format that expects the model to understand the nature of the task. This method uses the embedding of classes and words, associated to a vector format, thereby enabling the algorithm to recognize unrepresented text. This process is realized by associating text with other previously learned topics or words. Most state-of-the-art NLP models were trained for specific tasks, using supervised learning methods. However, these traditional Machine Learning models need a large amount of data to learn and fail to generalize tasks other than the ones they were trained on.

The architecture of GPT-2 is a 12-layer Decoder-only Transformer, trained within the *attention framework*³, rather than foregoing recurrence and convolutional-based architectures. GPT-2 is able to generate coherent (i.e. understandable by humans) paragraphs of text. The technology accomplishes state-of-the-art performance on language modelling, thereby outperforming benchmarks as RNN/LTSM/CNN-based models. GPT-2 performs basic reading comprehension, machine translation, Q&A and summarizing tasks. The model use cases range from writing grammar assistance and auto-completion (for ordinary and coding purposes), to more advanced applications such as conversational agents, creative writing engines, unsupervised translation models, and enhanced speech recognition

³ Attention mechanisms, as proposed by (Vaswani, 2017), enable the prediction model to discriminatively concentrate on sections of input text which are most important. Thus, attention mechanisms allow the modelling of dependencies regardless to their distance in the input or output sequence.

systems. The model was first released in stages using 124 million, 355 million, 774 million and finally 1.5 billion parameters, due to the authors concerns for malicious purposes, which could be generation of misleading news articles, impersonation of others online, automated production of abusive or faked content to post on social media and spam and phishing content (Radford et al. 2019).

2.2.2 GPT modules for task-oriented conversational agents

The majority of current dialog systems provide semantically correct responses; however, they strive to correspond to the variation and fluency in human language. NLG is the last and major phase to generate user-facing system speeches. The output of NLG is closely tied with the perceived quality and usability of a dialogue system (Zhu et al. 2019).

Current generative networks are trained on voluminous amounts of data providing them with the capacity to generalize to unseen sample (or test) data. However, once these models are deployed, there is no control over the generation. Although, transformer-based models have achieved good results on a range of NLP tasks in recent years, the lack of control remains a challenge.

To tackle the above problem, controllable text generation procedures have been established. The attributes to control can be of the following types (Prabhumoye et al. 2020):

- Stylistic Attributes such as politeness, sentiment, formality
- Demographic attributes such as gender or age
- Contextual attributes such as information, keywords, or entities
- Ordering of information such as events, e.g. plot summaries etc.

The majority of controllable text generation tasks can be framed as conditional language generation tasks and can be performed at various stages. Prabhumoye et al. (2020) divide the generation process into five modules, which can be used. As illustrated in figure 2, the generation pipeline starts with an input or a source sequence and an output or a target sequence. The first input module, i.e. the external input module, is responsible for the initialization h_0 . The second input module, i.e. is the sequential input module, denoted by x_t , where t is a time index of the generation process. The third module is the generator operations module. It is responsible for carrying out consistent operations or calculations on all the inputs at each time step. The fourth module is the output module. The output o_t is further projected on to the vocabulary space to predict the token $\widehat{X}_t$ at each time step. Finally, the fifth module is the training objective module. This module deals with the loss functions used for training.

There are various techniques used to modulate each of the five modules to perform controllable text generation sequences. In dialogue systems it would be meaningful to control the response generation

according to the semantic action. In other words, the response is conditioned on the dialogue act. For this purpose, the generator operation (module 3 in figure 2) should be considered. This module takes in the external input h_0 , the sequential input x_t at time step t and performs the same set of computations G to return an output o_t . Changes can be made to the set of operations G in order to include a so-called control vector in computing o_t .

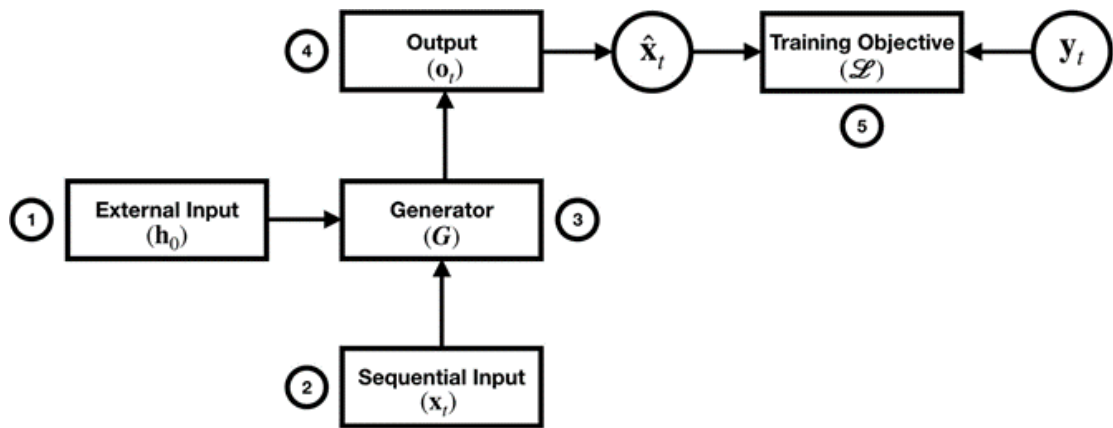


Figure 2: Modules for controlling a natural language generation task
Source: Prabhumoye et al. 2020

RNNs, LSTMs, and Gated Recurrent Units can be used to model controllable text generation tasks. However, these techniques still have trouble in handling long text sequences. As a result, attention mechanism algorithms are employed on the source sequence, as discussed in Prabhumoye et al. (2020). By incorporating self-attention, Transformers possess the capability to process input sequences of varying lengths and effectively capture the interdependencies among different slots. Peng et al. (2020) propose a multiple task and domain learning on large amounts of annotated (dialog act, response) pairs to the transformers pre-training stage. This optimization allows the developer to semantically condition the generation of responses of GPT models, and to adapt it to the task specific domain.

In order to improve the language model's behaviour, an iterative process named "Process for Adapting Language Models to Society (PALMS) with Values-Targeted Data-sets is used to fine-tune the model on a curated dataset. Empirical evidence shows that adjusted language models perform well even on small and hand-curated datasets. Additional tests conducted with GPT-3 even show considerably more robust results. The approach is promising and might be considered to further works with future GPT versions (Solaiman/Dennison 2021).

2.2.3 Our proposed hybrid solution for an IT-Support Q&A-System

In accordance with the internal guidelines for system development, the solution is to be developed as a micro-service architecture⁴, thus enabling re-usability and integrability with other applications. The proposed solution enables the end user to have a two-way text conversation with a conversational agent, without having to generate a ticket each time the users ask for assistance. Combining different technologies such as machine learning, NLP and integrated dialogue tools, our hybrid solution is able to maintain a conversational flow between applications and users.

The rules to trigger either the Rule-Based NLG module, or the GPT module, should be defined in advance. There are different possibilities for defining policies. One solution is to trigger GPT via a fallback intent. If no intent is found the conversational agent passes the user request to the GPT model. This would require GPT to be more generic, i.e. fine-tuned with several kinds of annotated response-request pairs to condition the answers, thereby following a dialog flow format. Another solution is to specifically define intents, which will trigger the GPT model to generate an answer. This alternative has nameable benefits, for instance a specific GPT model can be used and fine-tuned with response-request pairs extracted from the IT support's database. A predefined response text is not needed. Instead the output of GPT is passed (via conversational agent) as an answer.

After receiving user input through an application, it can be transmitted to the backend using, for example, a REST API. This standard procedure is frequently employed in enterprise settings, facilitating seamless integration of the system into existing frameworks and thus enabling expansion through in-store methods. The conversational agent plugin establishes communication with the application to retrieve the relevant input and obtain the corresponding output. Subsequent actions are carried out within the conversational agent plugin, and the resulting output is then sent back to the application. The conversational agent retains the context of the conversation and the user, enabling it to execute the necessary actions based on this information.

Through the conversation with the conversational agent, the user can receive faster responses compared to the traditional Information Technologies Service Management (ITSM) application structure. If the answers are not satisfactory, the user can ask the conversational agent to generate a ticket. This ticket will be assigned to the IT support and it will be identified with the user's name. The entire conversation history will be stored to support the request follow-up, allowing the subsequent IT specialist to understand the context of the underlying query. Thus, the ticketing system will be maintained. However, the interaction method should be changed from e-mail to conversational agent, which would perform the service at first level, trying to offer answers to FAQs. The user can access the conversa-

⁴ Micro-services approach to modularize software in systems which are divided into smaller modules. A micro-service should only be able to perform a single task and communicate with other programs. This enables micro-services to interact with each other (Wolff 2018).

tional agent application and directly ask a question, the conversational agent parses the natural language input, as provided by the end user, and triggers one of the NLG modules, according to the identified intent. This answers back with the solutions in the dialogue. Together with the given answer, a user feedback is asked, e.g. a question like “Was this answer helpful?” or “Was your question answered?”. If the feedback is negative (dashed line number 5 in figure 3), the conversational agent should escalate the request to second level support by triggering a creation of a ticket to the (human) IT-Support agent. Additionally, in case that the intent cannot be identified (a so-called *fallback*), the creation of a ticket to the IT-Support can also be triggered. This ticket can contain the conversation history, thereby ensuring that the technician is informed. In this phase the data for Q&R- pairs can be collected and used to enrich the IT service database. Note, that both negative, as well as positive, feedback can be considered, i.e. the first one indicating bad quality, or incorrect answers, and the second one indicating good quality, or correct answers. In this feasibility study our first focus lies on negative feedback loops to exclude early frustration potential on employees’ side. In future adoptions, positive feedback loops can be incorporated to increase the probability that the algorithm will give similar advices for similar requests. This represents a type of reinforcement learning strategy to improve our language model in next generations. The overall target will be to maximize user satisfaction.

Finally, the IT Support agent proceeds with the process by categorizing the ticket, working on it and replying to the end user. The communication channel, once a ticket is created, can follow the standard e-mailing process. If the conversational agent is unable to answer the user question, thus creating a ticket to the second-level IT Support, the user can receive both, a notification with the ticket number, as well as a response from the IT Support agent. Consequently, the only change on the current ticketing process is the conversational agent, which is creating tickets in name of the end user. The whole process is visualized in figure 3 once again.

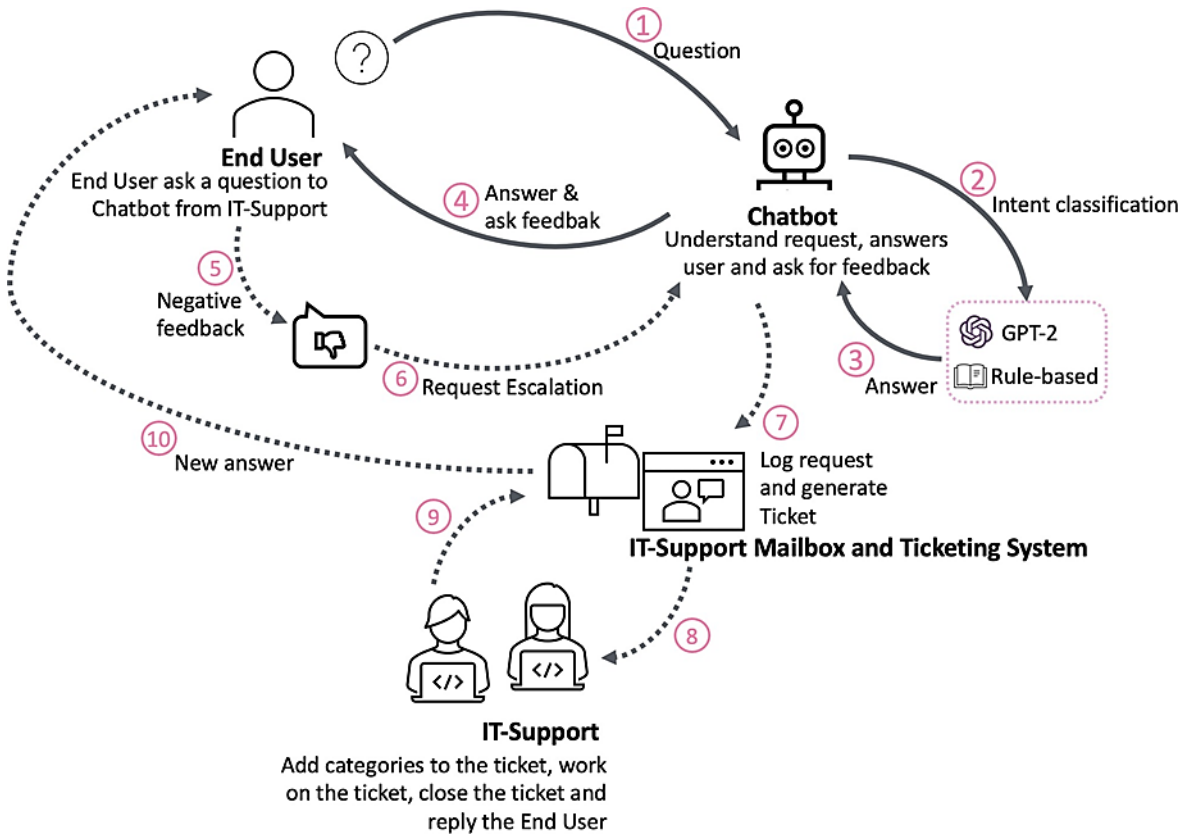


Figure 3: IT-Support process with a hybrid conversational agent for Q&A pairs
Source: authors' own illustration

Subsequently to the process implementation (including data collection for training purposes), the solution can evolve to a hybrid semantically conditioned conversational agent. The data from the initial phase can be used in an additional pre-training stage to semantically condition GPT generation; this would increase the quality of answers, as it would be more domain specific. Feedback loops (both positive and negative) should be continued.

3 A mini-workplan for implementing large-scale NLP models in IT Ticketing Systems

3.1 Towards developing a GPT-based System for Q&A tasks

This paper focuses on the IT support unit. Our approach aims to reduce the workload of a human IT specialist in order to work on larger scale IT problems. Moreover, faster replies increase the user's satisfaction, and vice versa, reduce opportunity costs resulting from ongoing conversation. We propose the use of a transformer-based NLP model, in this particular case, GPT. GPT is embedded in an

already structured IT ticketing process. We experiment with versions of GPT and further explore GPT's potential to work in a conversational agent solution for Q&A tasks. Our objective divides into the following stages:

- 1) Defining data sources and extraction methods: In order to enable our model to answer questions sent from the end users to the IT-Support, a sample of the data is needed. The first step is to evaluate how data is stored and how a sample can be extracted. Afterwards, pre-processing steps should be applied such as tokenization, removing numbers and extra white spaces.
- 2) Selecting a suitable GPT model and using few-shot learning methods on extracted data from the IT support: In this paper, we will make use of the Huggingface's repository to obtain a pre-trained GPT model. After selecting a model, the environment to perform the few-shot learning has to be configured to operate on the extracted sample from IT-Support.⁵
- 3) Compose a preliminary feasibility analysis of a Q&A System: We motivate the viability of using a hybrid solution. It aims to lead the tickets, which are currently assigned to human attendants, to an automated service. The idea is to illustrate that this filtering at the entrance and in the ticket ward will be able to promote the optimization of the resources used within the IT support unit. It further improves the quality of the service, thereby meeting the growing need of an instant communication tool to provide better user experience and to increases response capacity.

We propose to introduce GPT in the existing conversational agent model process. It is one of the most recent technologies for developing NLP models based upon transformers. GPT intends to take another step towards relating users and service providers. This approach allows the IT support to re-use resources, thus increasing ticket solution capacity.

3.2 The original IT Ticketing Workflow

The IT support unit consists of a team of specialists, who are the first point of contact for all internal and external users needing information technological guidance. If the user has not the ability to interact with the system, e.g. in form of a Q&A machine, the sole option is to generate a "ticket". That ticket is assigned in a queue until it is finally assigned to one of the attendants. This stage can take hours to days. A delay has severe consequences. Since it often prevents the user from continuing his or her work. However, in a variety of circumstances, the underlying problem belongs to a series of repeating questions that, if stored automatically, could have been resolved easily, and in a timely manner, by sending an automated reply emerging from a corpus of pre-trained solutions. The simplified process

⁵ The model as well as its checkpoints have to be imported. The checkpoint should be saved within "run1" folder since we are not training the model. Note, we only perform a few-shot learning algorithm. Further, the parameters of the model have to be defined, as well as its context with examples of Q&A pairs.

can be described as follows: The end user sends an email with a question (all end user requests, such as ordering Hardware/Software, Directory Services, Account configuration, granting Licenses/Subscriptions, Network Services, and more, will be referred to as Question) to the IT-Support mailbox. The Ticketing System, which runs in the background of this mailbox, identifies this email and generates a ticket associated to this specific user and question. This ticket is picked up by a member of the IT-Support team, who manually selects categories related to this question and starts working on the ticket. The end user has also access to the Ticketing System for tracking the status of the ticket (the status can be either open, in progress, or closed). When the ticket is resolved, IT support answers the end user and closes the ticket. This answer goes as an email response to the user. For sake of simplicity, only the 'happy flow' will be taken into account. Hence, scenarios where the IT-Support has to ask the end user for more information, tickets will be "escalated" to another IT Support agent. Further, the ticket has to be approved by a manager. Otherwise, the ticket cannot be marked as resolved. Figure 4 below illustrates the previously described process.

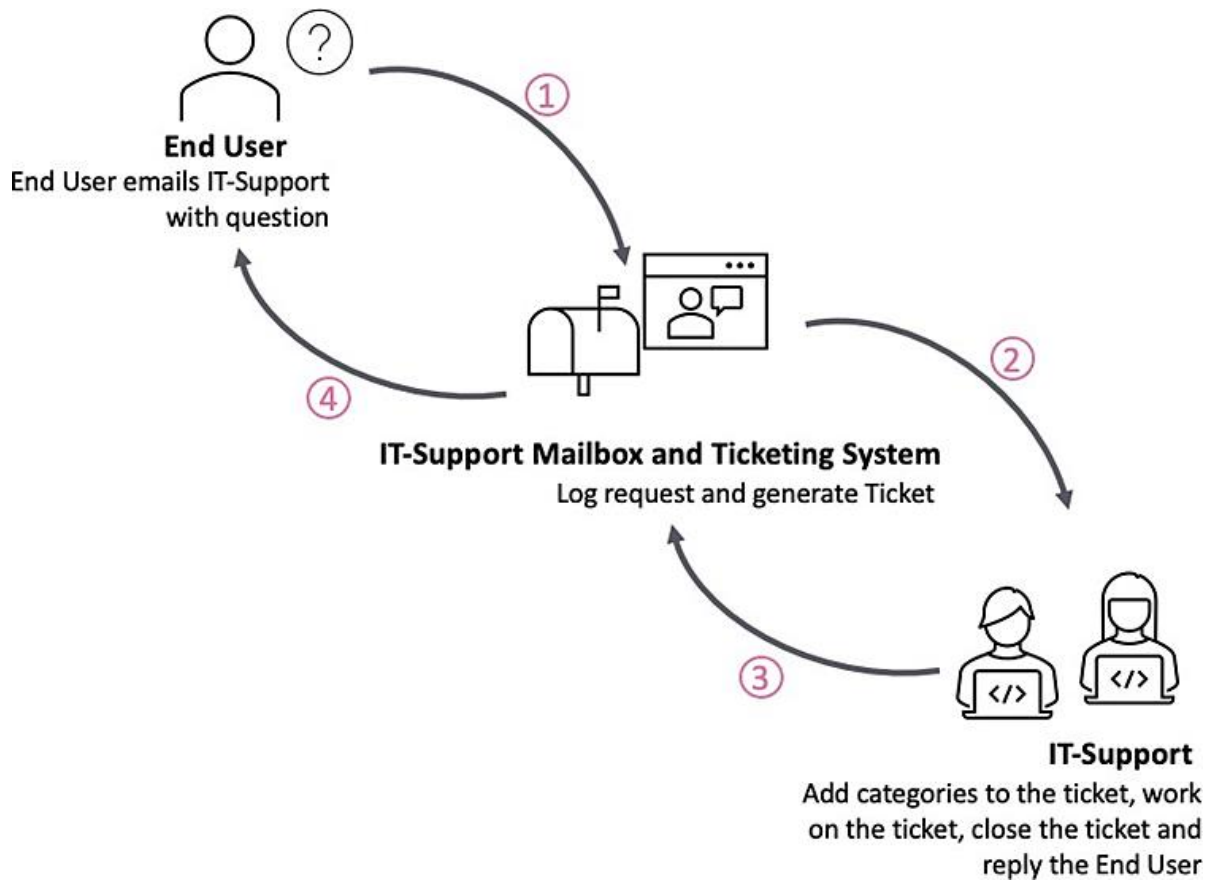


Figure 4: Q&A process within an IT ticketing system
Source: authors' own illustration

3.3 Data Gathering and Pre-processing

The data is stored in a semi-structured format in a MS-SQL-Database. The process is depicted from a data perspective in figure 5.

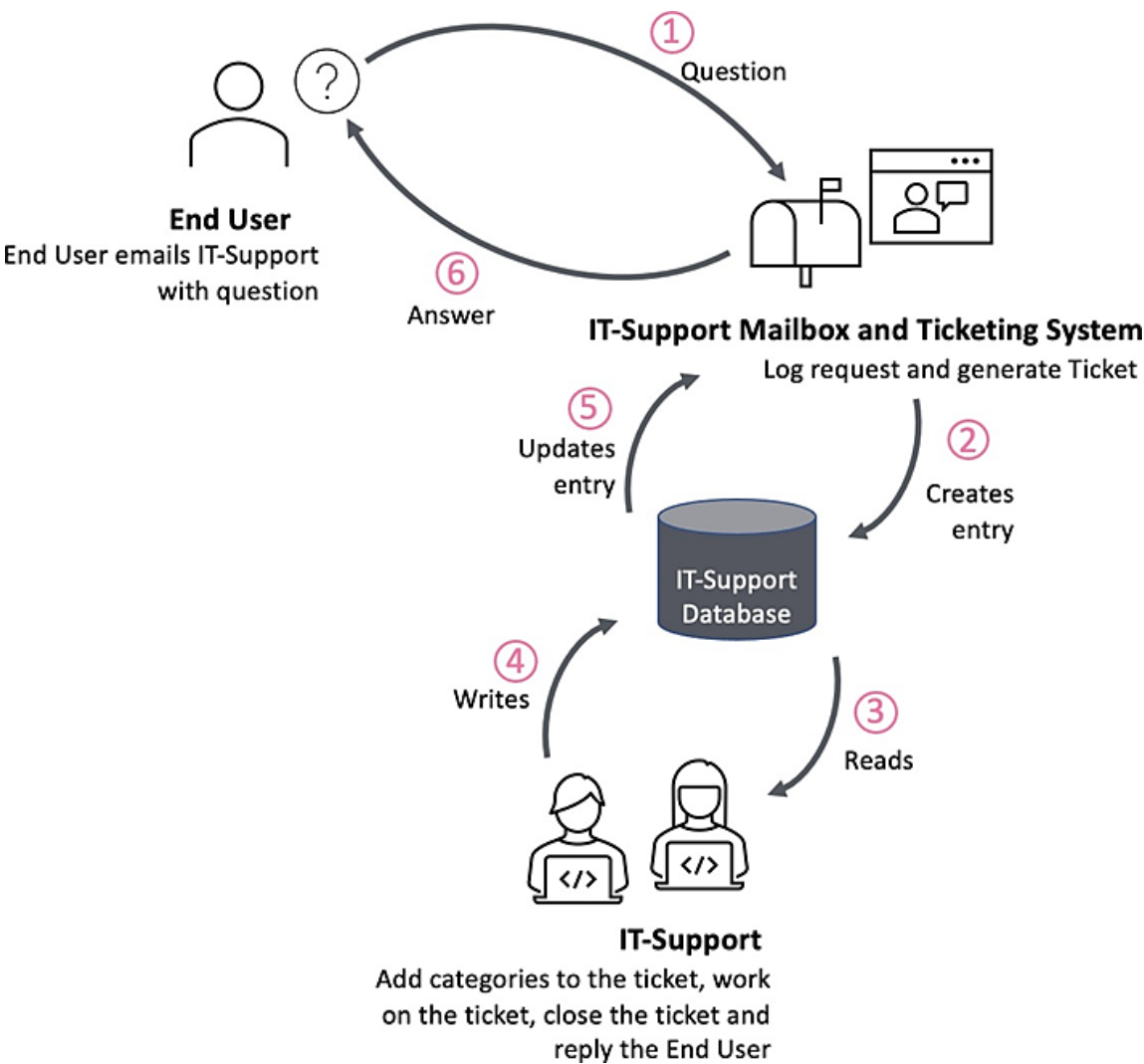


Figure 5: Exemplified data flow within a Q&A ticketing system
Source: authors' own illustration

The definition of data sources and extraction methods was conducted with the collaborating IT consultancy firm. Assessing data in a large-scale company demands considerable effort, due to restrictions, volume, and variety. As previously indicated, the data from the IT support is stored in a semi-structured format. All data is in German language. However, since most requests belong to an IT affiliated domain, most terms are in English, e.g. software names or licenses. Direct access to the database

was not granted due to the highly sensitive characteristics of the user data. In order to enable further analysis, i.e. on a “manageable” amount of data, a random sample with 733 MB of data and 76,508 observations, was generated.

To perform few-shot learning on the IT support dataset, a GPT model has to be configured to be able to answer in a dialog flow format. For this, training demonstrations are required that need additional pre-processing. Note that pre-processing is a very complex and time-consuming task in general. It is not the main topic of our paper. Therefore, only some basic procedures were adopted to meet the minimal viable criteria of feeding the model. The pre-processing was performed by Python by using the following libraries: pandas (for dealing with the tabular data frame), hashlib (for creation of cryptography hashes and message digests), and json (JavaScript Object Notation, a standardized format, such as XML and YAML, used to store and exchange data).

We have filtered only questions with less than 300 characters and related to License, tokenized the data by adding additional “Id”, “Request”, and “Response” columns and discarded all other columns. The pre-processed dataset was saved into a “json” file format with 352 lines, one for each observation that was inside the previously mentioned criteria.

4 Model Deployment and Evaluation

4.1 Objective Setting

The object of this paper, NLP for Q&A systems, will focus on the deployment within the current IT support process. We intend to present a hybrid solution, built upon a generative model which is able to lead the IT tickets, that are currently associated to human attendants, to an automated service.

Our analysis proposes the implementation of a conversational agent, for answering IT questions by using the GPT-2 technology. With this, a semi-structured component is brought to the predominantly human-orientated IT support. We believe that our solution has tremendous demand in software releases and licensing operations.

4.1.1 Task definition

In the following, we want to focus on IT questions related to software licenses. The choice was not only made, as first filtration in form of a feasibility study, but rather as an opportunity to automate an inherently structurable task. The contributing tasks are defined as follows:

1. Few-shot learning with *gpt-2-simple* and *german-gpt2*, using data from the IT support as few-shot demonstration;
2. Preliminary feasibility analysis for an IT support Q&A-System by using generative pre-trained transformers as underlying NLG module.

4.1.2 Technological decisions

For this study, several language models written in Python. The initial decision to go with a transformer-based language model was further refined into either an encoder only transformer, namely BERT (Bidirectional Encoder Representations from Transformers), or a decoder-only transformer model, specifically GPT-2 and GPT-3, by using the libraries *gpt-2-simple*, *german-gpt2* and *gpt3*.

The initial intention was to use an encoder-only transformers model, more specifically BERT (Bidirectional Encoder Representations from Transformers). BERT could be used to extract the answer from a span of text and provide it to the end user. The first problem with BERT is that the input is limited to 512 tokens (only). The second issue is that there is no available knowledge base for BERT to search its answer from. Therefore, using BERT on a large knowledge base would require 1) to find documents that are similar to the question, 2) to apply BERT to the most similar document, which would withdraw additional time resources from the company, 3) to use BERT as a classifier, to build FAQs from the past e-mails sent to the IT service desk. However, considering the given dataset, in combination with the computational restrictions, the GPT-2 (decoder-only transformers model) was assumed to generate more “tangible” results.⁶

The GPT-2 models used in this paper were taken from the Huggingface repository. All models within this repository are licensed under MIT⁷. For the programming framework, Google Colab⁸ was not a feasible solution, first of all due to data protection issues, secondly because the free version of Colab does not have sufficient TPU/GPU nor space (only 6 GB were available on Google drive) to train the model and save checkpoints. The models were implemented locally, on a Mac computer with 16 GB RAM, by using PyCharm⁹. PyCharm was used for programming in Python within Conda.¹⁰ Two virtual environments were created, one for each GPT model, with different Python interpreters in order to comply with the models' requirements.

⁶ We further tested and trained a model on the more recently developed GPT-3 counterpart, which is now commercialized but yielded significantly better results as we will see in section 4.3.

⁷ MIT license permits free of charge to any person with the copy of a software and associated documentation to deal with this software without restrictions to use, copy, modify, merge, publish, distribute, sub-license, and/or sell copies of the Software. For further disclaimer information see <https://mit-license.org>.

⁸ Colab is a web-browser cloud application used for writing and executing code in Python. Colab notebooks are Jupyter notebooks hosted by Colab.

⁹ PyCharm is an integrated development environment for programming in Python language.

¹⁰ Conda is a package and environment manager that can be used with command lines at the Windows prompt, or in a terminal window for Mac OS and Linux.

4.1.3 Pytorch vs. Tensorflow

In this topic we will assess a comparison between the Python Machine Learning frameworks of PyTorch and TensorFlow. On this paper both PyTorch and TensorFlow implementations of transformer models were tested.

Both frameworks run in a Python environment and have tools for quick visual analysis. PyTorch is built on Python classes. It allows the combination with other Python libraries, like SciPy and Numpy. Further, it extends the usual object functionalities (e.g. in printing statements at runtime by showing the data transferred from one object to another). In contrast to this, TensorFlow is a more advanced Machine Learning library, predominantly used for Deep Learning applications. Keras methods facilitate the use of TensorFlow. However, TensorFlow requires some experience with respect to functionality. PyTorch is not as feature-rich as TensorFlow, but all the essential features (necessary for our task) are available.¹¹

4.2 Few-shot learning with GPT-2

GPT-2 model is a language model trained on a wide variety of sources, allowing the model to incorporate idioms not seen in the input text. In other words, even though GPT-2 is not specifically trained on German language it should be able to generate text in German. A limiting factor might be the maximum number of tokens that can be generated by GPT-2, which amounts currently 1024. This corresponds to approximately two paragraphs of German text. But since we consider license requests, that have relatively straight forward requests and answers, we assume that the limiting number of tokens is sufficient for our language model.

Few-Shot (FS) will be used in this work as recommended by Brown et al. (2020). The model is given a few demonstrations of the task at hand, in this case IT tickets, and then conditioned (at some inference time) without updating the internal model weights. An example consists of a question and an answer (as desired completion). After being granted K examples of question and answers, in this case for license requests, in combination with one final example of a question for that the few-shot learner model is expected to provide an answer (completion task). To provide a vivid example of the question and answer pairs, we refer to figure 7.¹² The set of (training) examples with Q&A parts will be called context in the following.

¹¹ For further information see the *PyTorch* documentation.

¹² Although, this figure refers to the evaluation of the *test* set, it gives a good example of the questions in the *training* set since the train-test-split was chosen randomly, i.e. the Q&A parts shown in figure 7, could have also been used for the training set.

The main advantages of few-shot learners are 1) a major reduction in the need for task-specific data and 2) a reduced potential to learn an overly narrow distribution from a large but narrow fine-tuning dataset. The main disadvantage is that few-shot models are generally outperformed by state-of-the-art fine-tuned models.

The scope of this work is delimited to perform a few-shot demonstration with the pre-processed data from an IT-Support, built upon a GPT-2 model. To enable this task additional pre-processing steps are needed. Further, a “json” file containing Q&A pairs is created to shape the model's response into the direction of the intended response. This file served as context for the few-shot learning model. The overall process is again summarised in figure 6.

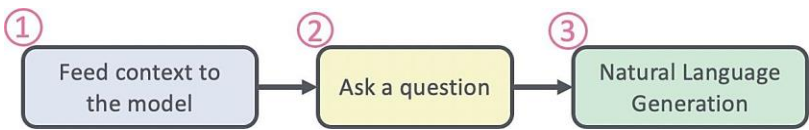


Figure 6: Few-shot learning mechanism with GPT-2
Source: authors' own illustration

4.2.1 Using the gpt-2-simple package

The Python package *gpt-2-simple* was initially developed by Max Woolf. It employs the *TensorFlow* library and combines fine-tuning and generating scripts for Open AI's text generation models.¹³ In order to use the *gpt-2-simple* package, the TensorFlow package (in many use cases) has to be downgraded to version 1.15. Furthermore, this version of TensorFlow is not supported anymore by Python 3.9 but the previous version of 3.7. The model only runs on the local system. The first release of GPT-2 with 117 M parameters, 12 layers of dimension 768 requires a noteworthy amount of disk space, i.e. 500 MB.

Further experimentation was performed by using checkpoints resulting into a medium model, with 345 M parameters and 24 layers of dimension 1,024, as well as a large model, with 774 M parameters and 36 layers of dimension 1,280. However, besides the increase in response time, accompanied by a non-negligible space consumption, the medium and large models did not result, in any significant improvement. In order to evaluate different models, we compared the generated answers with the labeled (i.e. true) answers provided by the IT support. For this, we used various text similarity (e.g. Cosine and Manhattan similarity).

¹³ For more details, see the *gpt-2-simple* package documentation.

4.2.2 Using the *german-gpt2* package

The *german-gpt-2* model is “pre-trained” with data provided by Bavarian State Library.¹⁴ The same corpora were used for the German BERT. In contrast to *gpt-2-simple*, this model uses the PyTorch implementation inside the transformer architecture of the Huggingface repository.

The model follows a similar process as *gpt-2-simple*. It uses few-shot learners as discussed in the previous section. An additional pre-processing step was added to allow the model evaluation, namely, a “demonstrations.txt” file, which uses the *json.loads()* method to parse a valid JSON string and converts it into a Python dictionary.

To run the model, a function was created, which takes the user question, the IT response, and the context, as parameters into account. The function calls the *text generation pipeline* (from Huggingface), which takes the task, the model, and the tokenizer (as input parameter and predicts the words that are most likely to follow. For our purpose, the task is set to *text-generation*, and the model as well as the tokenizer were set to *german-gpt2*.

4.3 Model Evaluation

Many NLG tasks seem to be open-ended. Academics, as well as industry practitioners, lead a constant debate for relevant evaluation metrics of generative models. However, the ultimate goal in NLG is to generate text, which is a) valuable and b) understandable for humans. For this precise reason, human judgments are a natural “first-choice” decision criterion for evaluating NLP models in many use cases (Celikyilmaz et al. 2020).

The evaluation process was performed by maintaining the same context. We manually insert a question for the model to answer. This allows us to compare answers of different models for one and the same question. It became evident that further pre-processing steps were necessary. Moreover, results showed that distance metric measures (such as the Euclidean distance) were impracticable for evaluating text similarities. More adequate text similarity measures were used instead such as the Levenshtein-ratio or the Cosine similarity. Finally, we decided to base our model validation on human judgment. For this, we compared the model responses for given questions to previous answers provided the human IT support. According to Hirschman and Gaizauskas (2001) model responses need to fulfill the following 6 criteria: Correctness, conciseness, completeness, coherence, relevance, and finally justification. It showed, that both *gpt-2 simple* and *german-gpt* models did not meet these expectations (see the appendix for an example of model predictions including the Levenshtein-ratio and the Cosine similarity). Further investigation made clear, that the lack of a question mark affected the

¹⁴ The MDZ Digital Library team from the Bavarian State Library (dbmdz for short) offers various open-source models within the Huggingface repository. For further information see the documentation.

model predictions. The repetition of one and the same phrase was another shortcoming that we identified especially for the *german-gpt2* model. The reason lies in the probabilistic model nature. Although *gpt2-simple* is not specifically pre-trained on German language, the model (partly) returns meaningful German phrases as the sample question in the appendix shows. However, the complex nature of Q&A tasks in general poses a major challenge for our model. This challenge is related to the inevitable need to formulate a question, as very open and unstructured text, which is still a difficult task to solve, even for modern NLP algorithms. To solve this problem, the engine needs additional knowledge about language structures, i.e. NLU capabilities as stated previous sections.

During the deployment of the previously discussed GPT-2 models, we were granted access right to the most recent GPT-3 API, called *Playground*. It provides a *general-purpose* “text-in-text-out” interface, which makes it applicable to more general language tasks. To use this API, the user can simply enter a text (prompt) in order to obtain completion. The present chosen task was “Q&A” and the parameters provided to the model were the same as for the *gpt2-simple* and the *german-gpt2* models. All other parameters of GPT-3 were left blank, thus indicating default values. For each model, *gpt2-simple*, *german-gpt2*, and *gpt3*, we followed the process depicted in figure 6 by using the same context for all Q&A training scenarios from license requests. The results, at least for GPT-3, were qualitatively stable, with respect to the above criteria posed by Hirschman and Gaizauskas (2001). An exemplary is shown in figure 7 (in German language). However, further experimentation and research is necessary in order to evaluate the capabilities and limitations of few-shot learner models for Q&A tasks in general. Model Limitations.

Although, we only worked on a small subsample for license requests, we are confident that our study generally shows the feasibility of a more general implementation including other IT relevant topics. We decided on the topic of “software licenses” since the underlying requests follow a clear structure, and moreover, in line with our anecdotal evidence from practice, already covers a large majority of daily IT requests. Being in the early stage of a feasibility analysis, this approach seemed to be the most plausible choice. If an easy structurable, task that poses a large majority of daily request, could not be covered by our model, we can have reasonable doubt to continue our task of implementing a Q&A system for IT help desks. We further believe that our large language model will improve with additional fine-tuning procedures. With enough training data, the model will also learn to build up more complicated, i.e. less structurable, dialog tasks. We seek to perform a structured hyperparameter-optimization (via grid-search), which we have refrained in our feasibility analysis.

GPT was chosen, as pre-trained NLP model. Besides its promising few-shot learning capabilities and actuality, the choice of a specific model poses a major challenge. This is due to inevitable links to BERT, which is an encoder-only transformer-based model for Q&A tasks.

Question:Hallo zusammen, hiermit beantrage ich eine IntelliJ Entwicklerlizenz für die Ultimate Edition.Viele Grüße Dennis

IT-Support Answer: Hallo Dennis, für die Beschaffung fallen Nettokosten in Höhe von insgesamt 499,00€ an. Bitte lass Dir die Kosten explizit von Deinem Vorgesetzten oder Deinem Projektleiter freigeben. In der Freigabe muss für uns ersichtlich sein, dass die Kosten dem Freigebenden bekannt sind. Das Ticket wird für 14 Tage zurückgestellt. Sofern die Freigabe innerhalb dieser 14 Tage bei uns eingeht, werden wir die Beschaffung durchführen. Wenn die Freigabe nach den 14 Tagen bei uns eingeht, wird für den kompletten Vorgang ein neues Ticket erstellt und wieder am Ende der Ticket-Queue einsortiert

gpt2-simple Answer: ich eine IntelliJ Entwicklerlizenz für die Ultimate Edition.BuyableInstoreAndOnline

german-gpt2 Answer: Hallo zusammen, ich möchte gerne nachfragen, ob wir Lizenzen für die Ultimate Edition haben? Ja, wir haben eine Lizenz für die Ultimate Edition. Einige Grüße Christian Flanders

GPT3 Answer: Hallo Dennis, für die Beschaffung fallen Nettokosten in Höhe von insgesamt 499,00€ an. Bitte lass Dir die Kosten explizit von Deinem Vorgesetzten oder Deinem Projektleiter freigeben. In der Freigabe muss für uns ersichtlich sein, dass die Kosten dem Freigebenden bekannt sind. as Ticket wird für 14 Tage zurückgestellt. Sofern die Freigabe innerhalb dieser 14 Tage bei uns eingeht, werden wir die Beschaffung durchführen. Wenn die Freigabe nach den 14 Tagen bei uns eingeht, wird für den kompletten Vorgang ein neues Ticket erstellt und wieder am Ende der Ticket-Queue einsortiert

Figure 7: Qualitative evaluation: Exemplary output of GPT-2-simple, German-gpt2 and GPT-3 for a request of an IntelliJ Developer license (in German language)

Moreover, the configuration of the virtual environment to perform the tests was not so simple. As mentioned before, many examples provided in the code documentation are made for using Google Colab notebooks that were inadequate for our tests (see section 4.1.2). Therefore, our models were stored locally. Lastly, the *gpt2-simple* model was finally selected for the sake of simplicity and model illustration purposes. With more time resources and budget, we would probably favour the most recently deployed GPT-3 standard. Fortunately, this decision can still be postponed until the deployment of the final Q&A ticketing system. In this case, even the adoption of the upcoming GPT-4 standard is even more desirable.

Additional problems arose from data quality. We found multiple Q&A pairs that were either not properly documented, or too complex and indirect. A further limitation is the quality of pre-processing steps. Each document needs to have a specific structure. The Q&A pair has to start with an identifying token. In addition to this, redundant information has to be removed. Both, the questions and answers, incorporated greetings, signatures, and other courtesies that are most likely to deteriorate our language model predictions.

Last but not least, the process of user feedback loops is key for our process and should be studied in more detail. The engagement of the corresponding IT support is also a crucial factor. The willingness to cooperate is characterized by a dichotomy between maintaining processes and increasing user experience.

4.4 Future research outlook

Micro-service architectures allow us to easily transfer currently developed NLP technologies to all other company domains. As outlined above, we propose to use conversational agents that fully rely on GPT technology. These agents can be transferred to all other business units that are exposed to constant communication exchange, e.g. “Human Resource Management” or “Administration Offices”. With enough “curated” dialog data even “Law and Compliance” solutions could benefit from NLP systems. Possible limitations might climax in generating strictly guided dialogues. However, a GPT-based solution represents a universal language model, without any need for intent classification. Consequently, beside the continuous improvement of NLP systems as a whole, future research should also concentrate on facilitating the implementation of such NLP systems in daily business operations.

5 Summary

In this study, we have designed an implementation guideline for setting up an NLP module to solve daily Q&A tasks for an internal IT help desk. Our language model, based on the recent GPT few-shot learning technologies such as GPT 2 and GPT 3. We have performed a feasibility study in cooperation with a large-scale IT consultancy company, at the example of software license requests. We have highlighted necessary technological objectives, i.e. the software packages to be used, as well as the data challenges to be overcome. Improvements result in a dynamic fine-tuning process, thereby increasing the training data support. The goal is to incorporate the module in the existing ticketing process, e.g. by using the in-store conversational agent solution. The module solves commonly structurable tasks. More complex IT tickets will be redistributed to human IT specialists as usual. However, the final deployment of our language model would result in a tremendous relief for the future workload of an IT help desk.

Appendix: Example question for evaluating GPT2 predictions for Q&A tasks in IT ticketing services including similarity scores

Question: Guten Morgen, ich brauche die Lizenz für die MS Tools und Microsoft Office. Vielen Dank! Beste Grüße DB Von meinem Samsung Galaxy Smartphone gesendet die Ultimate Edition.Viele Grüße Dennis

IT-Support Answer: Hallo Daniel, alle Office-Programme kannst Du unter portal.office.com auf bis zu 5 Geräten installieren. Benötigst Du darüber hinaus noch eine weitere Lizenz? Welche wäre das genau?

gpt2-simple Answer: Ich brauche die Lizenz für die MS Tools und Microsoft Office. Vielen Dank! Beste Grüße DB Von meinem Samsung Galaxy Smartphone gesendet.

Levenshtein-ratio: 0,955 **Cosine similarity:** 0,368.

german-gpt2 Answer: Hallo, ich brauche die Lizenz für die MS Tools und Microsoft Office.

Vielen Dank!

Levenshtein-ratio: 0,759 **Cosine similarity:** 0,282.

References

- Adamopoulou, E.; Moussiades, L. (2020): Chatbots. History, technology, and applications. In: *Machine Learning with Applications*, 2. (2020).
- Brown, T. B.; Mann, B.; Ryder, N.; Subbiah, M.; Kaplan, J.; Dhariwal, P. et al. (2020): Language Models are Few-Shot Learners. In: *arXiv preprint arXiv:2005.14165* (2020).
- Budzianowski, P.; Wen, T. H.; Tseng, B. H.; Casanueva, I.; Ultes, S.; Ramadan, O. & Gašić, M. (2018): Multiwoz— A large-scale multi-domain wizard-of-oz dataset for task-oriented dialogue modelling. In: *arXiv preprint arXiv:1810.00278*.
- Cheyner, A.; Didier Guzzoni, D. (2014): Method and apparatus for building an intelligent automated assistant. US Patent 8,677,377.
- Celikyilmaz, A.; Clark, E.; Gao, J. (2020): Evaluation of text generation. A survey. In: *arXiv preprint arXiv:2006.14799* (2020).
- Fiore, D.; Thiel, C.; Baldauf, M. (2020): Potenziale von Chatbots für den innerbetrieblichen IT-Support. In: *HMD Praxis der Wirtschaftsinformatik*, 57. (2020), No. 1, pp. 77–88.
- Hirschman, L.; Gaizauskas, R. (2001): Natural language question answering: the view from here. In: *natural language engineering*, 7. (2001), No. 4, pp. 275–300.
- Jäntti, M.; Cater-Steel, A.; Shrestha, A. (2012): Towards an improved it service desk system and processes. A Case Study. In: *International Journal on Advances in Systems and Measurements*, 5. (2012), No. 3 & 4, pp. 203–215.
- Kucherbaev, P.; Bozzon, A.; Houben, G.-J. (2018): Human-aided bots. In: *IEEE Internet Computing*, 22. (2018), No. 6, pp. 36–43.
- Langkilde I.; Knight, K. (1998): Generation that exploits corpus-based statistical knowledge. In *Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics-Volume 1*, pp. 704–710.
- Mathur, S.; Lopez, D. (2019): A scaled-down neural conversational model for chatbots. In: *Concurrency and Computation: Practice and Experience*, 31. (2019), No. 10.
- Peng, B.; Zhu, C.; Li, C.; Li, X.; Li, J.; Zeng, M. et al. (2020): Few-shot natural language generation for task-oriented dialog. In: *arXiv preprint arXiv:2002.12328* (2020).
- Prabhumoye, S.; Black, A. W.; Salakhutdinov, R. (2020): Exploring Controllable Text Generation Techniques. In: *arXiv preprint arXiv:2005.01822*.
- Radford, A.; Wu, J.; Amodei, D.; Amodei, D.; Clark, J.; Brundage, M. et al. (2019): Better language models and their implications (https://cdn.openai.com/better-language-models/language_models_are_unsupervised_multitask_learners.pdf). Accessed on 10/17/22.
- Solaiman, I.; Dennison, C. (2021): Process for Adapting Language Models to Society (PALMS) with Values-Targeted Datasets. In: *arXiv preprint arXiv:2106.10328* (2021).
- Tarlaci, F. (2019): GPT-2 for Question Answering (<https://fatmatarlaci.wordpress.com/2019/05/08/gpt2qa/>). Accessed on 05/25/21.

- Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A. N. et al. (2017): Attention is all you need. In: arXiv preprint arXiv:1706.03762 (2017).
- Vijayarani, S.; Ilamathi, M. J.; Nithya, M. et al. (2015): Preprocessing techniques for text mining-an overview. In: International Journal of Computer Science & Communication Networks, 5. (2015), No. 1, pp. 7–16.
- Wen, T. H.; Gasic, M.; Mrksic, N.; Su, P. H.; Vandyke, D. & Young, S. (2015): Semantically conditioned LSTM-based natural language generation for spoken dialogue systems. In: arXiv preprint arXiv:1508.01745.
- Wen, T. H.; Gasic, M.; Mrksic, N.; Rojas-Barahona, L. M.; Su, P. H.; Vandyke, D. & Young, S. (2016): Multi-domain neural network language generation for spoken dialogue systems. In: arXiv preprint arXiv:1603.01232.
- Wolff, E. (2018): Microservices. Grundlagen flexibler Softwarearchitekturen. 2.; aktual. Aufl.; Heidelberg: dpunkt.verlag.
- Zhu, C.; Zeng, M.; Huang, X. (2019): Multi-task Learning for Natural Language Generation in Task-Oriented Dialogue. In: Padó, S. (ed.): The 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing. Proceedings of system demonstrations. Conference on Empirical Methods in Natural Language Processing And the International Conference on Natural Language Processing (EMNLP-IJCNLP 2019), Hong Kong, China, November 3–7. Association for Computational Linguistics (ed.). Stroudsburg, PA : Association for Computational Linguistics (ACL), pp. 1261–1266. (<https://aclanthology.org/D19-1123.pdf>). Accessed on 04/05/23.

Authors

Prof. Dr. Marcus **Becker** is a full-time professor of “Quantitative Methods of Business Information Technology” at the International School of Management (ISM) in Dortmund. He gained international experience in automated Risk Management processes working for HSBC and Deloitte. Besides, he was part of the German Robo Advisory panel at Deloitte’s Risk Advisory service line testing Machine Learning algorithms like LSTM on stock market forecasts.



M. Sc. Erika **Prokop Dayrell de Lima**, is a former graduate student of ISM in “Business Intelligence and Data Science” (BIDS). She gained professional experience in data mining, machine learning, software engineering, and IT project management as a consultant working for Adesso SE. As a Requirements and Issue Manager at Kostal Group she is currently responsible for ensuring the implementation of requirements and resolution of issues throughout the development lifecycle.

