

GO>GROUP®

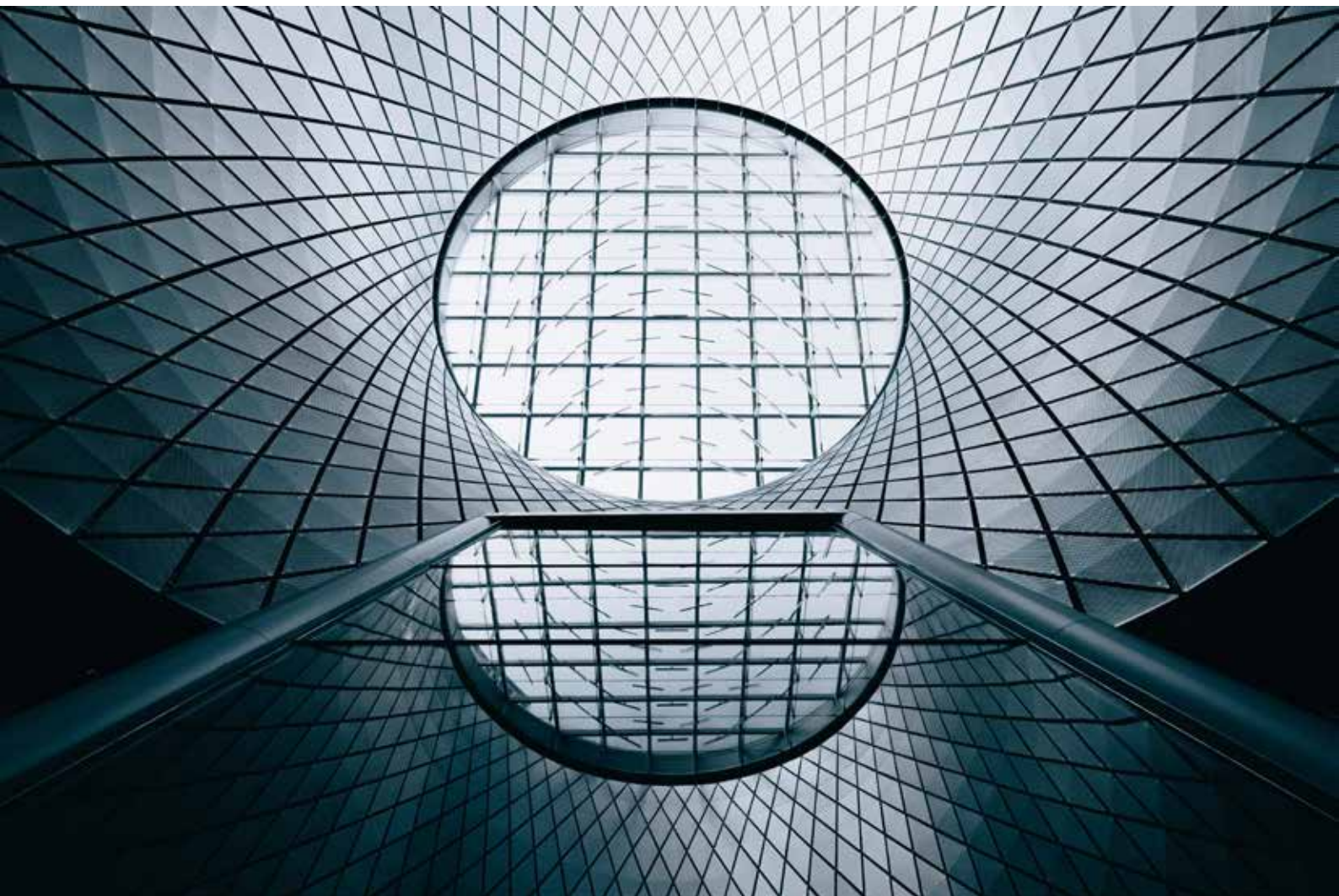
# Mustererkennung in der Optimierung

Meta-Analyse von A/B-Split-Tests in Deutschland



# VORWORT

Die vorliegende Studie ist eine Bachelor-Thesis, die in Zusammenarbeit mit unserem GO-Group Partner konversionsKRAFT in Bad Homburg entstanden ist.



# MANAGEMENT SUMMARY

Conversion Rate Optimization (CRO) using A / B split testing is an increasingly important process for e-business companies to maximize revenue without relying on increased website traffic. Especially for medium-sized companies whose traffic limits the number of possible split-tests, the effectiveness and efficiency of their CRO processes can be a decisive competitive advantage.

The purpose of this study is to identify factors that can be used to categorize and compare CRO processes to examine these factors for their effects, the likelihood of success, or the expected uplift of a split test.

The paper attempts to answer the following research questions: Are there any categorizable features of the CRO process, which are statistically dependent on the probability of success and the expected uplift of the split-tests performed, and if so, which? Are there statistical patterns among these characteristics that can serve as an indicator of the probability of success? Are there statistical relationships between several of these factors that allow predicting the probability of success?

For this purpose, a total of 448 conducted split tests of the company „Web Arts AG“ were statistically recorded, evaluated and analyzed. The sample data showed several characteristics and feature classes, with significant differences in success rates or uplifts achieved. As possible relevant features, the study could determine e.g. the type of page being tested or the psychological intent of the tested hypothesis. In addition, the work was able to provide a basis for statistical clustering and forecasting models based on the selected features.

On the basis of the research results, concrete recommendations for prioritization within a CRO process can already be derived. The results provide arguments for focusing on motivational hypotheses in CRO and tests for users on mobile devices. In further research, the investigation model can be extended by further variables or on the basis of further random samples, the developed approaches for a predictive model can be continued.

# ZUSAMMENFASSUNG

Die Conversion-Rate-Optimierung (CRO) mithilfe von A/B-Split-Tests ist für Unternehmen des E-Business ein stetig an Bedeutung gewinnender Prozess, um ihre Erträge zu erhöhen, ohne auf einen Zuwachs des Website-Traffics angewiesen zu sein. Als CRO-Prozess werden dabei alle Maßnahmen verstanden, die das Ziel haben die Conversion Rate zu erhöhen. Insbesondere für mittlere Unternehmen, deren Traffic die Anzahl der möglichen Split-Tests begrenzt, kann die Effektivität und Effizienz ihrer CRO-Prozesse ein entscheidender Wettbewerbsvorteil sein.

Diese Studie hat das Ziel, Faktoren zu bestimmen, mit denen sich CRO-Prozesse kategorisierbar und vergleichbar machen lassen, um diese Faktoren auf ihre Effekte, auf die Erfolgswahrscheinlichkeit oder den zu erwartenden Uplift eines Split-Tests zu untersuchen.

Die Arbeit soll folgende Forschungsfragen beantworten: Gibt es kategorisierbare Merkmale des CRO-Prozesses, zu denen die Erfolgswahrscheinlichkeit und der zu erwartende Uplift der durchgeführten Split-Tests in statistischer Abhängigkeit stehen, und wenn ja, welche? Gibt es unter diesen Merkmalen statistische Muster, die als Indikator für die Erfolgswahrscheinlichkeit dienen können? Gibt es statistische Beziehungen zwischen mehreren dieser Faktoren, die eine Vorhersage der Erfolgswahrscheinlichkeit erlauben?

Zu diesem Zweck wurden insgesamt 448 durchgeführte Split-Tests der Firma „Web Arts AG“ statistisch erfasst, bewertet und analysiert. Die Daten der Stichprobe zeigten mehrere Merkmale und Merkmalsklassen, mit signifikanten Unterschieden in den Erfolgsraten oder den erzielten Uplifts. Als mögliche relevante Merkmale konnte die Studie z.B. den Seitentyp, auf dem getestet wird, oder die psychologische Intention der getesteten Hypothese ermitteln. Zusätzlich konnte die Arbeit Grundlagen für statistische Cluster- und Vorhersagemodelle auf Basis der gewählten Merkmale schaffen.

Auf Basis der Forschungsergebnisse lassen sich bereits konkrete Empfehlungen für Priorisierungen innerhalb eines CRO-Prozesses ableiten. Die Ergebnisse liefern Argumente dafür, sich in der CRO auf motivationsgetriebene Hypothesen und auf Tests für Nutzer mit mobilen Endgeräten zu fokussieren. In der weiteren Forschung kann das Untersuchungsmodell um weitere Faktoren erweitert werden oder anhand weiterer Stichproben die erarbeiteten Ansätze für ein Vorhersagemodell fortgeführt werden.

# 01

# ÜBER DIESE ARBEIT

Die vorliegende Arbeit ist eine Bachelor-Thesis im Studienfach „E-Commerce“ und beschäftigt sich mit der Analyse von Prozessen hypothesenbasierten Split-Testings, zum Zwecke der Conversion-Rate-Optimierung (CRO) im E-Business, im Folgenden als „hypothesenbasiertes CRO“ oder „CRO-Prozesse“ bezeichnet.

In diesem Kapitel wird zunächst erklärt, was unter hypothesenbasiertem CRO verstanden wird und weshalb eine Erforschung dieser Prozesse relevant für die testenden Unternehmen ist. Anschließend werden die vorab gesetzten Ziele dieser Arbeit definiert und die zu diesem Zweck gewählte Vorgehensweise kurz beschrieben. Abschließend enthält das Kapitel eine kurze Beschreibung zum Aufbau der gesamten Arbeit und beschreibt die Umstände, unter denen diese zustande kam.

## 1.1 Warum Effektivität in der CRO so wichtig ist

Die vorliegende Arbeit ist eine Bachelor-Thesis im Studienfach „E-Commerce“ und beschäftigt sich mit der Analyse von Prozessen hypothesenbasierten Split-Testings, zum Zwecke der Conversion-Rate-Optimierung (CRO) im E-Business, im Folgenden als „hypothesenbasiertes CRO“ oder „CRO-Prozesse“ bezeichnet.

Die Conversion-Rate-Optimierung (CRO) ist eine Disziplin, die dem Online-Marketing zugezählt wird und die das Ziel verfolgt, auf Websites, Webapps oder mobilen Apps (im Folgenden immer einheitlich als Websites benannt) mehr erlösbringende Transaktionen aus der gleichen Anzahl an Besuchern zu gewinnen. Die Conversion Rate drückt dieses Verhältnis zwischen Transaktionen und der Anzahl von Website-Besuchern aus. (Vgl. Goward, 2013, S. 1)

In der Praxis geschieht diese Optimierung häufig in Form von hypothesengetriebenen A/B-Split-Tests, welche in iterativen Prozessen geplant, durchgeführt und analysiert werden. Die aufgestellten Hypothesen sollen erklären, welche Änderungen an einer Website die Nutzer dazu bewegt, häufiger zu konvertieren oder eine sonstige, vorab als Ziel definierte Handlung zu begehen. (Vgl. Reese, 2009, S. 20)

Die Ideen der Hypothese werden in einer oder mehreren Varianten der jeweiligen Seite umgesetzt und in einem überwachten Experiment mithilfe spezieller Testing-Tools, gegen die Original-Variante (V0) der Seite getestet. (vgl. Kauschik, 2010, S. 197) Durch die anschließende Auswertung der Ergebnisse erhält das testende Unternehmen nicht nur Daten darüber, ob die getestete Hypothese widerlegt wurde, sondern auch Erkenntnisse über das Verhalten der eigenen Nutzer, aus denen sich erneut Hypothesen ableiten lassen. (Vgl. Reese, 2009, S. 18–20)

Der Vorteil von CRO für das testende Unternehmen besteht darin, dass die Optimierungen an der eigenen Seite in der Regel für alle Besucher der Website gelten, unabhängig vom ursprünglichen Kanal, der die Besucher auf die Website gebracht hat. (vgl. Goward, 2013, S. 1) Viele Disziplinen des Online-Marketings, wie Suchmaschinen-, Social-Media- oder E-Mail-Marketing, dienen in erster Linie dem Zweck der Kundenakquise, also durch Investitionen zusätzlichen Traffic auf die eigene Website zu bekommen. (Vgl. Kreutzer, 2014, S. 2)





Die Ausgaben für die genannten Online-Marketing-Kanäle steigen (vgl. Deutsche Post AG, Juni 2018, S. 23), was die Effizienz der eingesetzten Mittel (Return on Advertising Spend oder ROAS) an Bedeutung gewinnen lässt. Für Unternehmen, deren Wachstumsmöglichkeiten durch eine steigende Marktsättigung beschränkt werden, ist die Steigerung der Produktivität, also der Conversion Rate, eine immer bedeutendere Kernaufgabe (vgl. Lauber, 2013, S. 106 f.) und sie kann effizientere Ergebnisse bringen als die Erschließung von neuen Nutzern durch Ausweitung der Marketing-Aktivitäten. (Vgl. Morys, 2011, S. 18 f.)

Viele Unternehmen erkennen CRO bereits als wichtigen Faktor für ihr Geschäftsmodell und betreiben CRO-Prozesse im eigenen Unternehmen oder lassen diese von externen Dienstleistern (Agenturen) durchführen (vgl. Bitkom e.V., Juli 2017, S. 24–27). In der Bitkom-Studie gaben 87% der befragten Unternehmen aus Deutschland, Österreich und der Schweiz an, dass der Anteil der Ausgaben für die Optimierung der eigenen Website oder des Webshops im kommenden Jahr steigen oder zumindest gleich hoch bleiben wird.

Auch international steigen die Ausgaben der Unternehmen in die CRO. In einer 2018 im Auf-trag des CRO-Tool-Anbieters „AB-Tasty“ durchgeführten Studie der „Forrester Research Inc.“ gaben 83% der befragten Unternehmen an, ihr Budget für CRO in den nächsten 12 Monaten um mindestens 10% erhöhen zu wollen. (vgl. Forrester Research Inc., 2018, S. 8) Laut einer Studie von „Econsultancy“ und „Adobe Systems“ von 2018 gaben 14% der 4.123 befragten Unternehmen und Agenturen an, CRO sei einer der drei digitalen Bereiche für ihre Organisation mit einer Top-Priorität im Jahr 2018. (Vgl. Econsultancy/Adobe Systems, März 2018, S. 23)

Die aufgeführten Studien belegen eine wachsende Bedeutung von CRO-Prozessen in Unternehmen. Eine Optimierung dieses Prozesses dient einer Verbesserung der Ertragssituation dieser Unternehmen, was unter der Maxime der Profitmaximierung eine betriebswirtschaftliche Notwendigkeit ist. (Vgl. Morys, 2011, S. 1)

Hat ein Unternehmen sich dazu entschieden, einen CRO-Prozess im eigenen Unternehmen zu implementieren, wird der Umfang der durchführbaren Split-Tests durch die Anzahl der Besucher der Website und die Anzahl der erzielten Conversions begrenzt, da jeder Test eine Mindestanzahl von Testteilnehmern und gemessenen Conversions benötigt, um statistisch belastbare (signifikante) Ergebnisse zu erhalten. (Vgl. Krüger, 2011, S. 17–19)

Da also, je nach Höhe des zum Testen verfügbaren Traffic, jeder durchgeführte Test einen eventuell bedeutenden Teil dieses „Traffic-Budgets“ verbraucht, sollten neben den eigentlichen Entwicklungskosten eines Split-Tests, die Opportunitätskosten für nicht durchführbare Tests als relevant erachtet werden.

Zweck dieser Arbeit ist es herauszufinden, welche Umstände dazu beitragen können, hypothesengetriebene CRO-Prozesse im Unternehmen erfolgreich zu gestalten, um das Betriebsergebnis im Allgemeinen zu verbessern und um im Speziellen sicherzustellen, dass das zum Testen zur Verfügung stehende „Traffic-Budget“ möglichst effektiv für erfolgversprechende Split-Tests verwendet wird.



## **Zielsetzung der Arbeit**

Das Ziel dieser Arbeit ist es, innere und äußere Faktoren einer prozessgetriebenen Conversion-Rate-Optimierung auszumachen, welche die Effektivität und die Effizienz dieser Prozesse sowohl positiv als auch negativ beeinflussen. Die Effektivität wird dabei als die Wahrscheinlichkeit bewertet, einen signifikant positiven Uplift mit mindestens einer getesteten Variante im Test zu erzielen. Die Effizienz bemisst sich an den im Durchschnitt zu erwartenden Uplifts der Conversion Rate bei einem erfolgreichen Test.

Zusätzlich soll überprüft werden, ob und welche bestimmten Konstellationen dieser betrachteten Faktoren im Zusammenhang mit der Erfolgswahrscheinlichkeit stehen. Beschrieben werden diese Konstellationen durch statistische Clustermodelle für „erfolgreiche“ und „nicht erfolgreiche“ Split-Tests.

Ein Schwerpunkt wird dabei auf die Betrachtung der zugrunde liegenden Hypothesen der durchgeführten Split-Tests gelegt. Diese werden im Sinne eines verhaltenspsychologischen Stimulus-Organismus-Response- (SOR-) Modells und des Behavior-Modells nach B. J. Fogg analysiert und bewertet.

Als Unterziel soll zu diesem Zweck zunächst ein Untersuchungsmodell für CRO-Prozesse entwickelt werden, mit dessen Hilfe eine Stichprobe mit Daten zu durchgeführten Split-Tests analysiert wird.

Die konkreten Forschungsfragen, welche die Arbeit beantworten möchte, leiten sich von der Zielsetzung ab und lauten wie folgt:

Forschungsfrage 1:

Gibt es kategorisierbare Merkmale des CRO-Prozesses, zu denen die Erfolgswahrscheinlichkeit und der zu erwartende Uplift der durchgeführten Split-Tests in statistischer Abhängigkeit stehen, und wenn ja, welche?

Forschungsfrage 2:

Gibt es unter diesen Merkmalen statistische Muster, die als Indikator für die Erfolgswahrscheinlichkeit dienen können?

Forschungsfrage 3:

Gibt es statistische Beziehungen zwischen mehreren dieser Faktoren, die eine Vorhersage der Erfolgswahrscheinlichkeit erlauben?

## **Aufbau der Arbeit**

Diese Arbeit ist in folgende fünf Abschnitte aufgeteilt:

Über diese Arbeit: Erklärt die Relevanz für die Erforschung der Effektivität von CRO-Prozessen und stellt die Ziele, Methoden und Umstände dieser Arbeit vor.

Theoretische Grundlagen der Untersuchung: Präsentiert ein Modell zur Abbildung des CRO-Prozesses und stellt Theorien aus der Verhaltens- und Konsumpsychologie vor, die das Zustandekommen ökonomischen Verhaltens erklären und damit die Grundlage bilden für die Auswahl der untersuchten Merkmale in der Studie.

Forschungsdesign: Beschreibt zunächst das Untersuchungsmodell und stellt alle unabhängigen und abhängigen Variablen vor. Danach wird das Vorgehen für die Beschaffung, Erfassung und Aufbereitung der untersuchten Stichproben-Daten erläutert. Schließlich wird erklärt, welche statistischen Methoden zur Beantwortung der Forschungsfragen benutzt wurden.

Auswertung der Studie und Beantwortung der Forschungsfragen: Präsentiert alle relevanten Ergebnisse der Analyse, hinterfragt die ermittelten Ergebnisse und formuliert Hypothesen zur Beantwortung der Forschungsfragen.

Fazit: Zum Schluss werden die wichtigsten Erkenntnisse der Studie zusammengefasst, bewertet und in einem unternehmensbezogenen Kontext betrachtet.

# 02 THEORETISCHER HINTERGRUND DER STUDIE

Diese Arbeit beschäftigt sich mit der Beschreibung und Untersuchung verschiedener Bereiche der Conversion-Rate-Optimierung. CRO wird häufig als eine Meta-Disziplin des Online-Marketings beschrieben, die Modelle, Methoden und Tools aus der Statistik, Psychologie, bestimmter Neurowissenschaften und des Marketings benutzt und darauf aufbaut. (Vgl. Morys, 2011, S. 37–40)

Die theoretischen Grundlagen für diese Arbeit lassen sich in drei Hauptbereiche einteilen. Zuerst wird der hypothesengetriebene CRO-Prozess modellhaft definiert. Danach wird mit Modellen der Psychologie und Neurowissenschaften erklärt, aus welcher Sichtweise die Hypothesen von Split-Tests in dieser Arbeit bewertet wurden. Anschließend wird mit dem 5C-Modell eine wissenschaftliche Einteilung der Geschäftsmodelle im E-Business vorgestellt, welche als Umfeldfaktoren eines CRO-Prozesses betrachtet wurden.

## 2.1 Modelle zur Beschreibung des CRO-Prozesses

Zu Beginn dieses Kapitels wird mit dem „DMAIC“-Zyklus ein Prozessmodell vorgestellt, welches ursprünglich in der Industrie für das Qualitätsmanagement entwickelt wurde und die Grundlage mehrerer Modelle für CRO-Prozessbeschreibungen ist. (vgl. Melzer, 2015, S. 5) Zusätzlich wird auf das „Conversion Cycle“-Modell von Nils Kattau im Speziellen eingegangen, um damit den Prozess einer hypothesengetriebenen Conversion-Rate-Optimierung modellhaft zu beschreiben.

### 2.1.1 Der DMAIC-Cycle als Referenzmodell für CRO-Prozesse

In der wissenschaftlichen Literatur findet sich kein Modell, das die speziellen Schritte eines hypothesengetriebenen CRO-Prozesses beschreibt. Jedoch finden sich mehrere Modelle, die von Experten aus der CRO-Praxis entwickelt und – aus unternehmerischen Interessen heraus – veröffentlicht wurden und die auf dem sogenannten „DMAIC“-Zyklus aufbauen, der ein etabliertes Modell zur Optimierung des Projekt- und Qualitätsmanagements ist. (Vgl. Melzer, 2015, S. 6)

Die Komponenten des „DMAIC“-Zyklus finden sich bereits in frühen Beschreibungen der Methodik in der CRO. Beispielsweise ist hier der „Usability Engineering Lifecycle“ zu nennen. (vgl. Beschnitt, 2009, S. 31) André Morys bezieht sich bei seiner Beschreibung des Conversion-Prozesses explizit auf den „DMAIC“-Zyklus. (Vgl. Morys, 2011, S. 202)

Der „DMAIC“-Zyklus gliedert den Prozess in fünf Phasen. Diese fünf Phasen in Reihenfolge sind:

Define, Measure, Analyze, Improve und Control. Die Phasen bauen aufeinander auf und für jede Phase gibt es eine Auswahl passender Werkzeuge und definierte Ergebnisse zum Abschluss der einzelnen Phase. (Vgl. Melzer, 2015, S. 4)

Im „DMAIC“-Zyklus dient die Define-Phase dazu, die Aufgabenstellung zu beschreiben, Ziele des Prozesses festzulegen sowie der Projektorganisation. In der Measure-Phase geht es um die Sammlung und Messung von Daten, um die Ist-Situation des Problems zu erfassen. Anschließend werden diese Daten in der Analyze-Phase dahingehend untersucht, Beziehungen zwischen signifikanten Ursachen (Inputs) und den entscheidenden Outputgrößen zu sammeln. Die Improve-Phase nutzt die Erkenntnisse und Ansätze aus der Analyze-Phase, um Maßnahmen zur Verbesserung zu entwickeln. Abschließend werden die Maßnahmen in der Control-Phase überwacht und gesteuert, um Effektivität und Nachhaltigkeit zu gewährleisten. (ebd.)

Morys weist noch speziell auf den iterativen Charakter des „DMAIC“-Zyklus in der CRO und die daraus entstehenden, kumulierenden Erkenntnisse jedes Prozess-Durchlaufs hin, welche in den folgenden Durchgängen genutzt werden können. (Vgl. Morys, 2011, S. 208)

### 2.1.2 Der Conversion Cycle zur Beschreibung des CRO-Prozesses

Ein aktuelles Modell zur Beschreibung hypothesengetriebener CRO-Prozesse ist der „Conversion Cycle“ von Nils Kattau, der in Abbildung 1 grafisch dargestellt ist. Im Gegensatz zu den zuvor erwähnten Modellen wurde der Conversion Cycle speziell für die Beschreibung von CRO-Prozessen entwickelt. Der Conversion Cycle beinhaltet die bewährten Aspekte des „DMAIC“-Zyklus und ergänzt diese um weitere Faktoren.

Wie in Abbildung 1 zu sehen ist, teilt der Conversion Cycle den CRO-Prozess in sieben Phasen auf. Diese werden von Kattau (2018 S. 81–85) folgendermaßen beschrieben:

**Analyse:** Identifiziert die Verbesserungspotenziale der Website

und liefert die Grundlagen zur Hypothesenbildung. Dies kann zum einen mit quantitativen Methoden geschehen, bspw. einem Webanalysetool wie Google Analytics, mit dessen Hilfe erkannt werden kann, wo, also auf welchen Seiten, und in welchem Umfang Probleme (bspw. viele Seitenabsprünge) auftreten.

Zum anderen kann in der Analysephase mit qualitativen Methoden erörtert werden, was genau die Schwachstellen an der jeweiligen Seite oder den Seiten sind. Als mögliche Methoden und Tools der qualitativen Analyse nennt Kattau Mouse-Tracking, Eye-Tracking-Simulationen, Conversion Frameworks, wie das „7-Ebenen“-Modell (vgl. Morys, 2011, S. 76) oder das „LIFT“-Modell (vgl. Goward, 09.03.2018), und den Einsatz von Wissen aus dem Neuromarketing und der Konsumpsychologie.



Abbildung 1 Conversion Cycle, Modell zur Abbildung eines CRO-Prozesses (Kattau, 2017, S. 81)

Aus dem Neuromarketing werden in der Praxis sogenannte kognitive Verzerrungen verwendet. (vgl. Wolf, 2018) Kognitive Verzerrungen sind psychologische Pattern und Motive, die im ökonomischen Kontext relevant sind und genutzt werden können. (Vgl. Beck, 2014, S. 9)

**Hypothese:** In dieser Phase werden, auf Basis der in der Analyse-Phase gewonnenen Erkenntnisse, Hypothesen darüber aufgestellt und gesammelt, welches beobachtete Verhalten durch welche Änderungen an der Website verändert werden soll. Eine mögliche Einschränkung auf bestimmte Nutzersegmente sowie Definitionen der messbaren Ziele zur Erfolgsbestimmung können ebenfalls Teil der Hypothesen sein.

Zusätzlich erfolgt in der Hypothesen-Phase noch eine Bewertung der Hypothesen, um diese nach Potenzial und Aufwand zu ordnen und zu priorisieren.

**Variation:** Die Ideen der Hypothesen werden in einer oder mehreren Test-Varianten umgesetzt. Teil dieser Phase sind auch vorbereitende Aufgaben, wie das Schreiben von Texten oder das Erstellen von Bildern oder Grafiken, welche anschließend in der Test-Variante verwendet werden. Das Ergebnis der Variation-Phase kann ein Mock-Up der Variante/n sein oder eine im Browser aufrufbare HTML-Version.

**Technical Setup:** In der technischen Umsetzung werden die zu testenden Varianten inkl. der Original-Variante mithilfe eines A/B-Testing-Tools für das Web umgesetzt.

Das Testing-Tool übernimmt dabei die Stichprobenauswahl sowie die technische Ausspielung der Test-Varianten und bietet

Einstellungsmöglichkeiten zur Segmentierung des Traffics.

Ein weiterer relevanter Aspekt dieser Prozessphase ist die Qualitätssicherung. Diese soll sicherstellen, dass die erstellten Varianten auf allen im Test eingeschlossenen Endgeräten (bspw. Desktop-Monitore bis zu einer gewissen Größe und Smartphones) ohne auftretende Artefakte dargestellt werden, die Funktionsweise der Variante/n gewährleistet und keine ungewollten Abweichungen zur Original-Variante vorhanden sind, welche nicht Teil der zugrunde liegenden Hypothese waren. (Vgl. Kwan, 2018)

**Monitoring:** Der eigentliche Split-Test wird während der Monitoring-Phase durchgeführt. Mithilfe des Testing-Tools und/oder weiterer Webanalysetools wird die Performance der Varianten in Bezug auf die eingestellten Ziele und Unterziele gemessen. Der Test kann nach Erreichen folgender Kriterien abgeschlossen werden:

- Das gemessene Testresultat erreicht ein festgelegtes Signifikanzniveau alpha. In der Regel gilt  $\alpha = 5\%$
- Testlaufzeit: mind. eine Woche
- Vorgegebene Mindestanzahl an Conversions je Variante
- Möglichst parallellaufende Graphen der kumulativen Conversion Rates von Original- und „Gewinner“-Variante

**Interpretation:** In dieser Phase werden die Testresultate analysiert und bewertet. Die Testergebnisse werden neben den reinen Leistungsdaten in Bezug auf die primären Conversion-Ziele, insbesondere auch auf direkte und indirekte Erkenntnisse hin überprüft, welche als Analysegrundlage für weitere Hypothesen dienen können.



Ebenso erfolgt eine Dokumentation des Testvorgangs inklusive der Hypothese, Abbildungen und Beschreibung der getesteten Varianten sowie den Testergebnissen.

**Implementierung:** Wenn die Testvariante im Split-Test eine signifikant höhere Conversion Rate erzielen konnte als die Original-Variante, wird diese in der abschließenden Implementierungs-Phase dauerhaft im Quellcode der Website umgesetzt.

Ein nächster Durchgang des „Conversion Cycle“ beginnt entweder erneut in der Analyse-Phase oder setzt eine zuvor erstellte und nicht umgesetzte Hypothese ab der Variation-Phase um.

Der „Conversion Cycle“ ist ein spezielles Modell für die CRO, ohne den für Projekte allgemein anwendbaren Ansatz des „DMAIC“-Zyklus. (vgl. Melzer, 2015, S.3) Durch die Spezialisierung stellt das Modell die praktische Realität in der CRO detaillierter dar als der „DMAIC“-Zyklus und beinhaltet dennoch alle wichtigen Elemente von eben diesem. Die Spezialisierung des Modells von Kattau und dessen Aktualität waren ausschlaggebend dafür, den „Conversion Cycle“ als Grundlage für den Betrachtungsraum des Untersuchungsmodells dieser Studie zu verwenden.

## 2.2. Verhaltenspsychologische Modelle zur Herbeiführung einer Verhaltensänderung

Wie im vorangegangenen Unterkapitel beschrieben, erklärt eine Test-Hypothese in der CRO, durch welche Veränderungen an der Website das Verhalten der Nutzer geändert bzw. in eine bestimmte Richtung gelenkt werden soll. Der „Conversion Cycle“ zeigt ebenfalls, dass die Hypothese, wenngleich sie nicht das alleinige Erfolgskriterium im CRO-Prozess ist, dennoch eine zentrale Rolle innerhalb des Zyklus innehat.

Während die Analyse-Phase dazu dient, Daten und Erkenntnisse zu gewinnen, von denen sich Hypothesen ableiten lassen, bauen alle weiteren Phasen ab der Variation-Phase auf der ausgewählten Hypothese auf, setzen diese in Test-Varianten um und bewerten die daraus resultierenden Testergebnisse.

Aus diesem Grund wurde beim Studiendesign ein Fokus auf die Erfassung und Beschreibung der Test-Hypothesen gelegt. Als Grundlage für die Beschreibung der Hypothese stützt die Studie sich auf Modelle aus der Konsum- und Verhaltenspsychologie, die beschreiben, wie eine Verhaltensänderung im ökonomischen Umfeld durch äußere Einflüsse herbeigeführt werden kann.

In den folgenden Abschnitten wird zunächst auf Stimulus-Response-(SR-) und Stimulus-Organismus-Response-(SOR-)Modelle eingegangen, welche die zusammenwirkenden Mechanismen von äußeren Reizen auf den Organismus und den daraus resultierenden Handlungen erklären.

Das anschließend vorgestellte „Behavior“-Modell beschreibt, welche psychologischen Trigger innerhalb des Organismus das Verhalten beeinflussen.

### 2.2.1 SR- und SOR-Modelle der Verhaltenspsychologie

Der Behaviorismus wie auch der Neobehaviorismus beschäftigen sich mit der Wirkung von messbaren äußeren Reizen auf die messbare Veränderung im Verhalten des Reizempfängers. (vgl. Kroeber-Riel/Gröppel-Klein, 2013, S. 55) Während das behavioristische Forschungsparadigma, als SR-Modell bezeichnet, lediglich den „Input“-Faktor Reiz bzw. Stimulus und den daraus resultierenden „Output“ (Verhalten bzw. Response) betrachtet, wird im Neobehaviorismus zusätzlich erforscht, wie die äußeren Reize auf das Innere des Reizempfängers wirken. (Vgl. Foscht/Swoboda/Schramm-Klein, 2015, S. 23)

Die SR-Modelle werden dazu um die Ebene des Organismus erweitert und bilden sogenannte SOR-Modelle. Abbildung 2 zeigt alle Variablen und Wirkungsrichtungen des SOR-Modells.

Alle Variablen des SOR-Modells sind in zwei Klassen eingeteilt: beobachtbare und nicht beobachtbare Variablen, sogenannte intervenierende Variablen. Um intervenierende Variablen messbar zu machen, werden messbare oder zumindest beobachtbare Phänomene (Indikatoren) benötigt. (Ebd.)

Stimuli, die einen Reiz auf den Organismus ausüben, werden in Marketing- und Umfeld-Stimuli aufgeteilt. Unter Umfeld-Stimuli werden äußere Umstände verstanden, die den Bedarf eines Produkts oder einer Dienstleistung entstehen lassen oder verstärken, bspw. eine gesetzliche Pflicht zur Mitführung von Warnwesten in einem PKW.



Abbildung 2 Prinzipiendarstellung des SOR-Modells (Foscht/Swoboda/Schramm-Klein, 2015, S. 30)

Marketing-Stimuli stammen in der Regel vom Hersteller oder Verkäufer. Darunter fallen bspw. die Produktpreise oder Werbekommunikation im klassischen Sinn. Auch eine Kommunikation mittels Website fällt unter diese Kategorien und daher kann eine Optimierung dieser Website als Stimulus im Sinne eines SOR-Modells verstanden werden.

Neben den beobachtbaren Stimuli enthält das Modell prädisponierende Faktoren, bspw. die kulturelle Prägung eines Menschen, welche ebenfalls Einfluss auf den zentralen Organismus und das Verhalten haben. (Vgl. Foscht/Swoboda/Schramm-Klein, 2015, S. 30)

Im SOR-Modell werden die inneren (psychischen) Vorgänge grundsätzlich in aktivierende und kognitive Prozesse unterteilt. Als aktivierend werden solche Vorgänge bezeichnet, die mit inneren Erregungen und Spannungen verbunden sind und das Verhalten antreiben. Kognitiv sind solche Vorgänge, durch die das Individuum Informationen aufnimmt, verarbeitet und speichert. (Vgl. Kroeber-Riel/Gröppel-Klein, 2013, S. 51)

Die aktivierenden Vorgänge teilen sich wiederum in die spezifischen Antriebe „Emotion“, „Motivation“ und „Einstellung“ auf. Während Foscht u. a. in ihrer Vorstellung des Modells, wie sie in Abbildung 2 zu sehen ist, die Einstellung des Konsumenten als eine Art Meta-Ebene im Organismus verstehen, bewerten Kroeber-Riel und Gröppel-Klein die Einstellung als einen aktivierenden Prozess.

Das daraus resultierende Verhalten (Response) ist dann wieder messbar, bspw. als Anzahl von Käufen oder dem durchschnittlichen Bestellwert.

Der Ablauf des SOR-Modells findet sich im Untersuchungsmodell dieser Studie wieder, indem zu den getesteten Hypothesen ihre jeweilig verwendeten Stimuli, deren Wirkung auf den Organismus und die gewünschte Reaktion kategorisiert und erfasst wurden.

Optimierungen an einer Website stützen sich in der Praxis jedoch nicht nur auf konsumpsychologische Muster, sondern verfolgen häufig auch das Ziel, technische Hürden bei der Benutzung einer Website zu beheben, bspw. durch eine Verbesserung von Ladezeiten oder eine nutzerfreundlichere Bedienbarkeit (Usability). (Vgl. Krüger, 2011, S. 85)

Da diese Intention im SOR-Modell nicht trennscharf zuzuordnen ist, verwendete die Studie zur Klassifizierung der Vorgänge im Organismus die Variablen des Behavior-Modells, welches im folgenden Abschnitt vorgestellt wird.

### 2.2.2 Das Behavior-Modell

Foggs-Behavior-Modell (FBM) wurde von B. J. Fogg an der Universität von Stanford entwickelt. Es stuft die Vorgänge innerhalb des Organismus ein, die zu einer ökonomischen Verhaltensänderung führen können. (Vgl. Fogg, 2009, S. 1–5)

Nach dem Modell ist das menschliche Verhalten – und damit auch das Konsumentenverhalten – das Produkt aus den drei Faktoren Motivation, Ability und Trigger, welche im Organismus zusammen- und wechselwirken. Die drei Faktoren bestehen wiederum aus Unterkategorien und werden zum besseren Verständnis einzeln vorgestellt:

**Motivation:** Diese liegt jeder Handlung zugrunde. Ohne Motivation wird keine Handlung erfolgen. Die Motivation für ein bestimmtes Verhalten, bspw. den Kauf eines Konsumgutes in einem Online-Shop, kann individuell stark ausgeprägt sein und kann durch äußere Faktoren beeinflusst werden.

In der Motivationspsychologie wird die Motivation in Anreizklassen unterteilt, welche die der Motivation zugrunde liegenden Beweggründe klassifizieren. Die am häufigsten untersuchten Anreizklassen (Motivthemen) sind die Leistungsmotivation (der Wunsch, eine Herausforderung zu meistern), die Anschlussmotivation (der Wunsch nach sozialem Kontakt) und die Machtmotivation (der Wunsch, andere Menschen zu beeinflussen oder zu beeindrucken). (Vgl. Brandstätter u. a., 2018, S. 5)

Bei Fogg bildet sich die Motivation aus den drei sogenannten Motivatoren „Pleasure/Pain“, „Hope/Fear“ und „Social Acceptance/Rejection“. (vgl. Fogg, 2009, S. 4)

Vergnügen und Schmerz im Sinne des Modells, zeichnen sich durch ihre sofortige Wirkung aus und lassen sich laut Fogg auf triebische und instinktive Bedürfnisse wie Hunger und Sex adaptieren. Hoffnung und Angst zeichnen sich durch die Erwartungshaltung an ein bestimmtes Ergebnis aus. Hoffnung steht für die Erwartung von etwas Positivem, während Angst bedeutet, etwas Negatives zu erwarten, häufig z. B. einen Verlust. Die Dimension sozialer Akzeptanz und Zurückweisung bestimmt laut Fogg einen Großteil unserer sozialen Handlungen.

Beide Ansätze zur Erklärung der Motivation, sowohl die Unterteilung in Anreizklassen, wie Brandstätter sie beschreibt, als auch die Teilung in Motivatoren bei Fogg, legen dar, dass die tatsächliche Motivation sich meist aus einer Kombination der zugrunde liegenden Motivationstreiber bildet.

**Ability:** Beschreibt die Fähigkeit zum Handeln bzw. wie einfach oder schwer ein gewünschtes Verhalten ausgeübt werden kann. Fogg nennt sechs Arten der Ability: Zeit, Geld, physische Anstrengung, kognitive Anstrengung, soziale Abweichung und Routine. Vier dieser Formen lassen sich der Disziplin Usability zuordnen. Eine gute Usability befähigt den Nutzer zeitlich, kognitiv und physisch zur Handlung, ebenso wenn die Handlung bereits routiniert ist.

**Trigger:** Sind die eigentlichen Auslöser einer Handlung. Trigger fordern ihren Empfänger direkt zu einer bestimmten Handlung auf. Fogg nennt drei Arten von Triggern – Sparks, Facilitators und Signals –, die sich in ihrer Funktion unterscheiden.



Abbildung 3 Signal-Trigger am Beispiel einer Twitter-Menü-Leiste.

Sparks sind mit motivierenden Elementen verbunden und dienen deswegen dazu, die Motivation zu erhöhen. Dies kann bspw. durch eine prominente Platzierung eines Produktvideos geschehen oder durch einen Call-to-Action-(CTA-)Button mit motivierendem Inhalt („Jetzt kaufen und sparen“).

Facilitators werden in Situationen angewendet, bei denen der Nutzer zwar eine hohe Motivation hat, aber ein Defizit in der Ability vorliegt. Ziel dieses Trigger-Typs ist es, zur gewünschten Handlung aufzufordern und diese gleichzeitig einfacher zu machen oder zumindest einfacher erscheinen zu lassen.

Fogg nennt als Beispiel die Funktion einiger Social-Media-Apps für das Smartphone, mit der alle gespeicherten Kontakte des Smartphones durch einen Klick mit dem Social-Media-Profil zu verknüpft werden können. Facilitators können auch als abstrakte Trigger funktionieren, etwa eine

Layout-Anpassung, um die Aufmerksamkeit des Nutzers auf eine bestimmte Funktion zu lenken.

Die dritte Art von Triggern im FBM sind Signale, die quasi als Erinnerung zur Durchführung der Handlung dienen, wenn Motivation und Ability ausreichend vorhanden sind. Signale motivieren nicht und vereinfachen auch keine Handlung. Sie lassen sich mit dem grünen Signallicht einer Verkehrsampel vergleichen, das den Verkehrsteilnehmer lediglich darauf hinweist, jetzt losfahren zu können. Als Signale auf Websites im Sinne des FBM werden zum Beispiel ein einfacher CTA-Button mit der Aufforderung „in den Warenkorb“ verstanden oder die grafischen Hinweise in einem Social-Media-Profil, dass neue Nachrichten vorhanden sind (siehe Abbildung 3). Der Screenshot aus Abbildung 3 zeigt eine Twitter Menüleiste. Die „5“ auf blauem Hintergrund, dient dem Nutzer als Signal dazu, die neuen Mitteilungen zu öffnen.

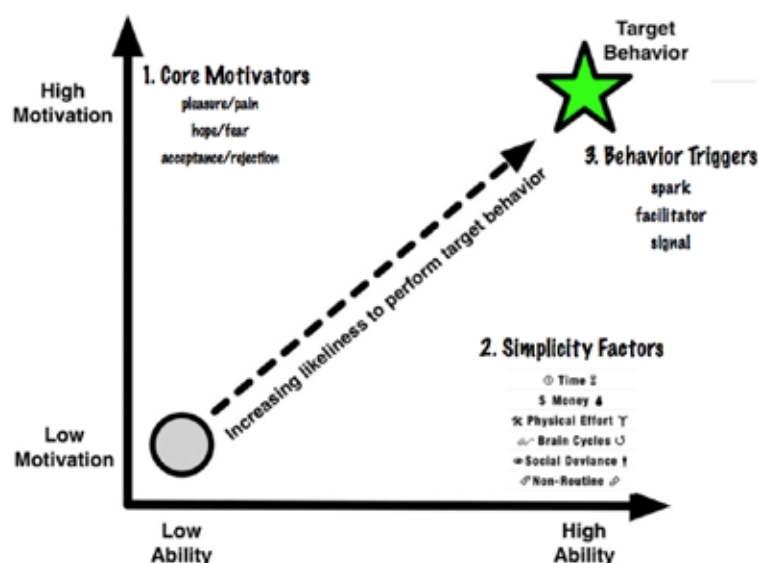


Abbildung 4 Visualisierung des FBM mit den 3 Hauptkategorien sowie den jeweiligen Unterklassen (Fogg, 2009, 5)

Die drei Faktoren Motivation, Ability und Trigger sind im FBM in ihrer Reihenfolge nach Bedeutung geordnet. Die Motivation ist Grundvoraussetzung für jedes Handeln, Ability ermöglicht oder erschwert dieses. Bei nicht vorhandener oder zu geringer Motivation wird auch eine aus-geprägte Fähigkeit nicht zu einer Handlung führen. Trigger müssen sich in Art und Intensität an die Ausprägungen der anderen beiden Faktoren anpassen, um zur Handlung zu bewegen. Abbildung 4 zeigt eine mögliche grafische Darstellung des FBM, welche das Spannungsverhältnis der drei Hauptkategorien des Modells veranschaulicht.

Die drei Hauptfaktoren des FBM dienen in dieser Studie dazu, die Intention/en einer Hypothese zu kategorisieren. In der Praxis ist die tatsächliche Intention zumeist ein Produkt von mehreren oder allen Faktoren des Modells. (vgl. Morys, 2011, S. 72) Weswegen auch in der Studie die Zuweisung von zwei oder drei Intentionen möglich war.

### **2.3 Weitere mögliche Einflussfaktoren für den Erfolg eines Split-Tests**

Da es sich bei Split-Tests um Feldexperimente handelt, können Umfeld und Bedingungen, in denen der Test durchgeführt wurde, wie bei jedem wissenschaftlichen Experiment stets Einfluss auf das beobachtete Ergebnis haben. (Vgl. Balzert u. a., 2010, S. 67)

Deswegen wurden neben den bisher vorgestellten Modellen, welche die einzelnen Arbeitsschritte und Methoden des CRO-Prozess definieren oder den Typ und die Intention des CRO-Experiments klassifizieren sollen, auch der Kontext und das Umfeld, in dem ein Split-Test durchgeführt wurde, in die Betrachtungsebene der Arbeit mit aufgenommen.

Das Umfeld, in dem ein Split-Test durchgeführt wird, ist u. a. das E-Business-Unternehmen, auf dessen Website getestet wird.

„E-Business ist die Nutzung von innovativen Informationstechnologien für die Vorbereitung [...], Verhandlung [...], und Durchführung [...] von Geschäftsprozessen zwischen ökonomischen Partnern über innovative Kommunikationsnetzwerke [das Internet und das Web, Anm. d. A.]“ (Kollmann, 2011, S. 55) .

Die unterschiedlichen Geschäftsfelder von E-Business-Unternehmen definieren sich über die Art der Wertschöpfung, den Erlös-Typ (vgl. Zerdick, 2001, S. 25) und die Akteure innerhalb des Prozesses. (Vgl. Wirtz, 2016, 30)

Wirtz unterscheidet zwischen den Geschäftsfeldern E-Commerce, E-Collaboration, E-Communication, E-Education und E-Information/Entertainment . Im Modell wird der Begriff der E-Information/Entertainment jedoch relativ breit gefasst, wodurch Infotainment-Content-Anbieter (bspw. Nachrichtenseiten) gleichgestellt würden mit Content-Seiten, deren Ziel eine Weitervermittlung (bspw. Affiliate-Seiten) oder Leadgenerierung von Kunden ist. (Vgl. Wirtz, 2016, S. 34)

Da bereits früh abzusehen war, dass eine relevante Anzahl der untersuchten Experimente von Websites stammen, deren primäres Ziel die Leadgenerierung und -vermittlung ist, stützt diese Arbeit sich auf ein Modell von Kollmann (2011), welches die genannten Geschäftsmodelle in sich berücksichtigt.



Kollmanns Modell ist ein sogenanntes 5C-Modell, das sich so und in vergleichbarer Form öfter in der Literatur findet. (vgl. Martínez-López, 2014, S. 4–9) Die Geschäftskonzepte des E-Business werden im 5C-Modell aufgeteilt in:

**Content:** Beinhaltet die Sammlung, Selektion, Systematisierung, Kompilierung und Bereitstellung von Inhalten gegen Bezahlung oder zur freien Bereitstellung mit indirekten Erlösmodellen, etwa durch Werbeeinnahmen. Beispiele des Geschäftskonzepts sind Nachrichtenseiten wie „spiegelonline.de“ oder „theguardian.co.uk“.

**Commerce:** Umfasst die Anbahnung, Aushandlung bzw. Abwicklung von geschäftlichen Transaktionen, bspw. Online-Shops.

**Connection:** Ermöglicht und organisiert die technologische oder kommerzielle Interaktion von Akteuren in Datennetzen. Neben einer technischen „Connection“, wie sie von Internet Providern bereitgestellt wird, nennt Kollmann als ergänzendes Beispiel für eine kommerzielle Zusammenführung „autoscout24.de“, eine Website, die Autohändler zum Zwecke des Gebrauchtwagenverkaufs mit einer Datenbankanbindung auf einen elektronischen Marktplatz bringt.

Connection: Ermöglicht und organisiert die technologische oder kommerzielle Interaktion von Akteuren in Datennetzen. Neben einer technischen „Connection“, wie sie von Internet Providern bereitgestellt wird, nennt Kollmann als ergänzendes Beispiel für eine kommerzielle Zusammenführung „autoscout24.de“, eine Website, die Autohändler zum Zwecke des

Gebrauchtwagenverkaufs mit einer Datenbankanbindung auf einen elektronischen Marktplatz bringt.

**Context:** Das Geschäftskonzept erbringt eine Klassifizierung, Systematisierung und Zusammenführung von verfügbaren Informationen und Leistungen in Netzwerken, mit den Zielen einer Komplexitätsreduktion und einer verbesserten Orientierung für ihre Nutzer. In erster Linie sind hier Suchsysteme wie „google“ oder auch „yelp“ zu nennen.

**Communication:** Ermöglicht und unterstützt die Interaktion, insbesondere die Kommunikation zwischen Akteuren in Netzwerken. Beispiele hierfür sind elektronische Communities wie das soziale Netzwerk „facebook“ oder die Dating-App „tinder“.

Die Art des zugrunde liegenden Geschäftskonzepts stellt den Website-Betreiber vor unterschiedliche Anforderungen (vgl. Kollmann, 2011, S. 60–62) und auch in der Optimierung werden für jeden Geschäftstyp, unterschiedliche Ansätze gewählt. (Vgl. Ash/Ginty/Page, 2012, S. 57)

# 03 DATENMODELL

Das Vorgehen zur Erstellung dieser Studie teilte sich in drei Hauptabschnitte. Diese Phasen und die darin verwendeten Methoden werden im folgenden Kapitel beschrieben. Zuerst wurde, auf Basis der in Kapitel 2 vorgestellten theoretischen Grundlagen, ein Forschungsmodell entwickelt, dass alle zu erhebenden Variablen zusammenfasst und sinnvoll ordnet. Anschließend wurden die Daten für die Stichprobe erhoben und aufbereitet. In der abschließenden Auswertung, wurden geeignete statistische Methoden und Formeln gewählt um die Stichprobe zu analysieren und die Ergebnisse zu erhalten, welche für die Beantwortung der Forschungsfragen und das Erreichen der Ziele dieser Arbeit notwendig waren.

### 3.1 Die Klassifizierung des CRO-Prozesses mithilfe des Untersuchungsmodells

Um die Ziele der Arbeit zu erreichen, musste zunächst entschieden werden, welche Faktoren in die Untersuchung miteinbezogen werden sollten. Alle zu erfassenden Variablen wurden in einem Forschungsmodell zusammengefasst. Abbildung 5 zeigt eine grafische Zusammenfassung des gewählten Forschungsmodells.

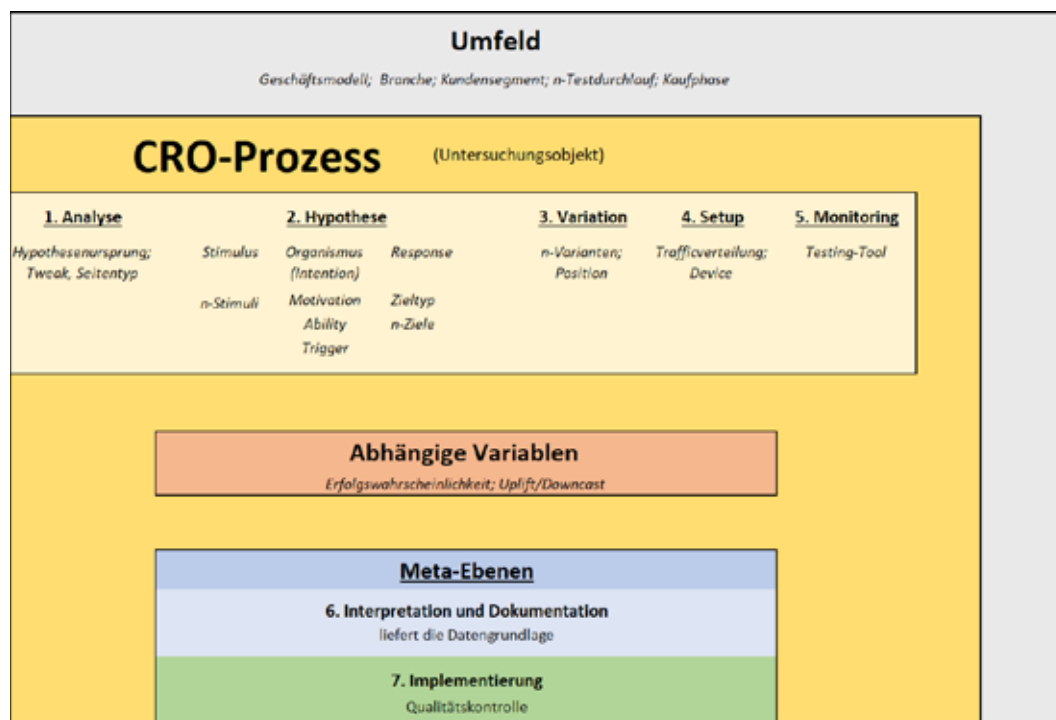


Abbildung 5 Grafische Darstellung des angewandten Untersuchungsmodells

Die ausgewählten Faktoren ergeben sich aus den vorgestellten Modellen für den CRO-Prozess, den psychologischen Theorien des SOR- und Behavior-Modells und einigen Umfeld-faktoren, etwa der Einteilung in E-Business-Geschäftsmodelle als Umfeldfaktor des CRO-Prozesses, wie diese in Kapitel 2.3 vorgestellt wurden.

Wie in Abbildung 5 zu sehen ist, wurden die erfassten Variablen in einzelne Bereiche aufgeteilt. Die Aufteilung erfolgte durch eine Zuweisung zu den Umfeldfaktoren oder in die einzelnen Phasen des „Conversion Cycle“, der in Kapitel 2.1.2 ausführlich vorgestellt wurde.

Die Phasen 6 und 7 im CRO-Prozess dienen im Modell als Meta-Ebenen. Die durchgeführten Tätigkeiten aus Phase 6, also eine ausreichende Dokumentation der Split-Tests, lieferte die Datengrundlage dafür, die Tests nachträglich erfassen und kategorisieren zu können. Phase 7 des Conversion Cycle, die Implementierung der Gewinner-Varianten, ist eine hypothetische Qualitätskontrolle im Modell. Durch eine dauerhafte Erfassung und Dokumentation der tatsächlichen Conversion Rate des testenden Unternehmens, können die im Experiment erfassten Uplifts validiert und damit die Testergebnisse bewertet werden.

Um die angewandte Logik und Interpretation bei der Datenerfassung nachvollziehen zu können, werden in Tabelle 1 alle unabhängigen Variablen aufgezählt und kurz im Sinne der Studie interpretiert und/oder durch Aufzählung aller möglichen Merkmalsausprägungen definiert.

Variable	Beschreibung
<b>Umfeldfaktoren</b>	
Branche	Branche oder Kernsortiment (bei E-Commerce Seiten) der testenden Website. Bsp.: Urlaub & Reisen; Finanzen & Versicherungen; Konsumgüter.
Geschäftsmodell	E-Business-Aktivitäten der testenden Website nach Kollmann, wie in Kapitel 2.3.1 vorgestellt. Ergebnisraum {Commerce; Connection; Communication; Content; Collaboration}
Kundensegment	Ergebnisraum {B2B; B2C}
n-Testdurchlauf	= Anzahl der bisher durchgeführten Tests an einer Website + 1
<b>CRO-Prozess</b>	
Hypothesen-Ursprung	Methode oder Framework, welches zur Erstellung der Hypothese genutzt wurde. Beispiele: Tweak, Persona, Conversion Framework (Lift, 7-Ebenen-Modell), User-Lab
Seitentyp	Ergebnisraum {Startseite; Produkt-Landingpage; Kategorieseite/n; Global; ... }
Stimulus/i	Art der getesteten Änderung in VbestVariante. Was wurde auf der Website geändert oder hinzugefügt? Beispiele: Wording, Bilder, UVPs, Design, Websitestruktur.
Intention	Ergebnisraum {Motivation (M); Ability (A); Trigger (T); MA; MT; MAT; AT}
Conversion-Typ	Ergebnisraum {Macro-Quantitativ; Macro-Qualitativ; Micro-Quantitativ; Micro-Qualitativ }
n-Ziele	Anzahl der gemessenen Ziele während der Testlaufzeit
n-Varianten	Anzahl aller getesteten Varianten im Split-Test
Position	Bereich der einzelnen Seite, in welchem die primären Änderungen des Testkonzepts durchgeführt wurden. Ergebnisraum {Header, Above the fold, Below the fold, Sidebar, Footer, Main, Relaunch}
Testing-Tool	Tool, mit dem der Split-Test technisch realisiert wurde. Beispiele: Optimizely, Google Optimize
Device	Ergebnisraum {Mobile; Desktop; Responsive}

Tabelle 1 Übersicht aller unabhängigen Variablen der Untersuchung



### 3.2 Vorgehen und Methodik in der Datenanalyse

Um die unterschiedlichen Forschungsfragen beantworten zu können, wurden passende statistische Methoden zu jeder Fragestellung gewählt. Die praktische Berechnung und Analyse wurden mit Funktionen der Computerprogramme „Microsoft Excel 2016“ und „Klime Analytics Platform“ durchgeführt.

Um den Zusammenhang einzelner Merkmale mit der Erfolgswahrscheinlichkeit zu untersuchen, wurden die Werte jeder Merkmalsausprägung und deren Häufigkeiten für erfolgreiche und nicht erfolgreiche Tests in einer Kontingenztafel zusammengefasst. Mithilfe des Chi<sup>2</sup>-Unabhängigkeitstests wurde anschließend analysiert, ob die einzelnen Merkmalsausprägungen im Zusammenhang mit der Erfolgswahrscheinlichkeit stehen.

Der Chi<sup>2</sup>-Unabhängigkeitstest vergleicht dabei die beobachteten Werte der aufgestellten Kontingenztafel mit den erwarteten Werten, die bei einer Unabhängigkeit der Faktoren im Durchschnitt auftreten sollten. (Vgl. Fahrmeir u. a., 2003, S. 464–466)

Um Zusammenhänge der einzelnen Faktoren mit den im Durchschnitt erzielten Uplifts abzuleiten, wurden die jeweiligen Stichproben mit einem Zwei-Stichproben-t-Test analysiert.

Mit dem t-Test wird die Verteilung der Werte um den Mittelwert einer mindestens metrisch skalierten abhängigen Variablen in zwei Stichproben verglichen. Als Ergebnis erhält man die statistische Wahrscheinlichkeit, dass beide Stichproben sich tatsächlich unterscheiden. (Vgl. Kähler, 2002, S. 364–366)

Um Zusammenhänge zwischen den Faktoren untereinander zu untersuchen und diese in Abhängigkeit des Testergebnisses zu setzen, wurde jeweils eine k-Means-Clustering-Analyse für alle erfolgreichen Tests durchgeführt und eine für alle nicht erfolgreichen Tests.

Die Clusteranalyse ist ein Verfahren der Mustererkennung in Stichproben mit mehreren beobachteten Faktoren. Sie ermöglicht, die Stichproben in eine bestimmte Menge „k“-Cluster aufzuteilen, deren Einheiten untereinander deutlich größere Ähnlichkeit besitzen als zwischen den verschiedenen Clustern. (Vgl. Wiedenbeck/Züll, 2010, S. 524 f.)



Durch die Klassifizierung in erfolgreiche und nicht erfolgreiche Tests können die jeweiligen Cluster verglichen werden. Durch unterschiedliche Cluster in beiden Stichproben sollen Hypothesen über bestimmte Muster für erfolgreiche und nicht erfolgreiche Tests abgeleitet werden.

Um den Einfluss mehrerer Faktoren auf die kategoriale Zielvariable „Erfolg“ zu untersuchen und daraus ein Vorhersagemodell zu erstellen, eignet sich eine Regressionsanalyse mittels logistischer Regression. (Vgl. Fahrmeir u. a., 2003, S. 504)

Da es sich bei den Prädiktoren (unabhängige Variablen) der Regressionsanalyse ebenfalls um kategoriale Merkmale handelt, mussten diese zunächst aufgelöst werden. Dies geschah durch eine Dummy-Kodierung mit den Parametern 1 als Indikator für das Auftreten einer zugewiesenen Merkmalsausprägung und 0 als Indikator für das Fehlen bzw. Nichtauftreten der jeweiligen Merkmalsausprägung. (Vgl. Rohrlack, 2007, S. 269)

Die Berechnungen des Clustering-Algorithmus sowie der logistischen Regression wurden mit Hilfe zweier Analyse-Workflows in dem Computerprogramm „Knime Analytics Platform“ durchgeführt. Eine Darstellung der verwendeten Workflows und Screenshots der verwendeten Einstellungen im Programm stehen im Anhang dieser Arbeit in den Abbildungen 10-14 auf den Seiten 50 und folgende.



# 04 AUSWERTUNG

Das folgende Kapitel präsentiert die relevanten Ergebnisse der durchgeführten Studie, insbesondere jene, die zur Beantwortung der Forschungsfragen nötig waren. Zuerst werden die Hauptkennzahlen und die Zusammensetzung der Stichprobe präsentiert. Anschließend werden die Forschungsfragen nacheinander anhand der jeweils relevanten Studienergebnisse beantwortet.

#### 4.1 Zusammensetzung der Stichprobe

In der Studie wurden insgesamt (N=) 448 unterschiedliche Experimente untersucht. Davon waren 187 erfolgreich, also mit einem signifikant positiven Uplift auf das primäre Ziel. Dies entspricht einer Erfolgsquote von 41,7% aller Tests. (vgl. Abbildung 6) Die 187 erfolgreichen Tests der Stichprobe bildeten die Grundgesamtheit aller Tests für die Untersuchungen, den Uplift betreffend. Abbildung 7 zeigt die Verteilung der erzielten Uplifts aus diesen 187 erfolgreichen Testergebnissen in einem Boxplot-Diagramm.



Abbildung 6 Erfolgsrate innerhalb der Stichprobe

Die gemessenen Uplifts der Stichprobe liegen zwischen dem Minimalwert von 0,5 und einem Maximalwert von 141,9. Das Boxplot-Diagramm in Abbildung 7 visualisiert die Verteilung der erzielten Uplifts. Innerhalb des Bereichs der blauen Box in Abbildung 7 liegen die mittleren 50% aller Werte der Stichprobe. Diese liegen um den Medianwert 9,5 und variieren zwischen 5,2 und 17,9.

Der Mittelwert der Stichprobe beträgt 16,1 und die Standardabweichung +/-20,6. Die Werte der äußeren Antennen des Boxplots stehen für den letztniedrigsten und letzthöchsten Wert der Stichprobe, dessen Lage noch innerhalb der Standardabweichung um den Mittelwert liegt.

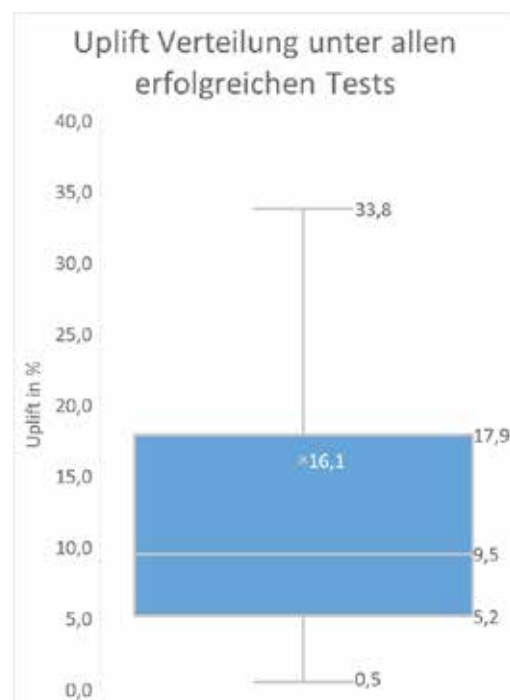


Abbildung 7 Verteilung der erzielten Uplifts unter allen erfolgreichen Tests

Die Experimente stammen von 56 unterschiedlichen Websites. Wie in Abbildung 8 zu sehen ist, ist die Verteilung der untersuchten Tests auf die einzelnen Websites nicht gleichmäßig. Abbildung 8 Verteilung der Stichprobe auf einzelne Websites

Den Daten aus Abbildung 8 ist zu entnehmen, dass 31% aller untersuchten Tests, von 3 Websites stammen. Knapp die Hälfte aller analysierten Tests stammen von neun Websites, während die andere Hälfte aller Tests, von 47 unterschiedlichen Websites stammen.

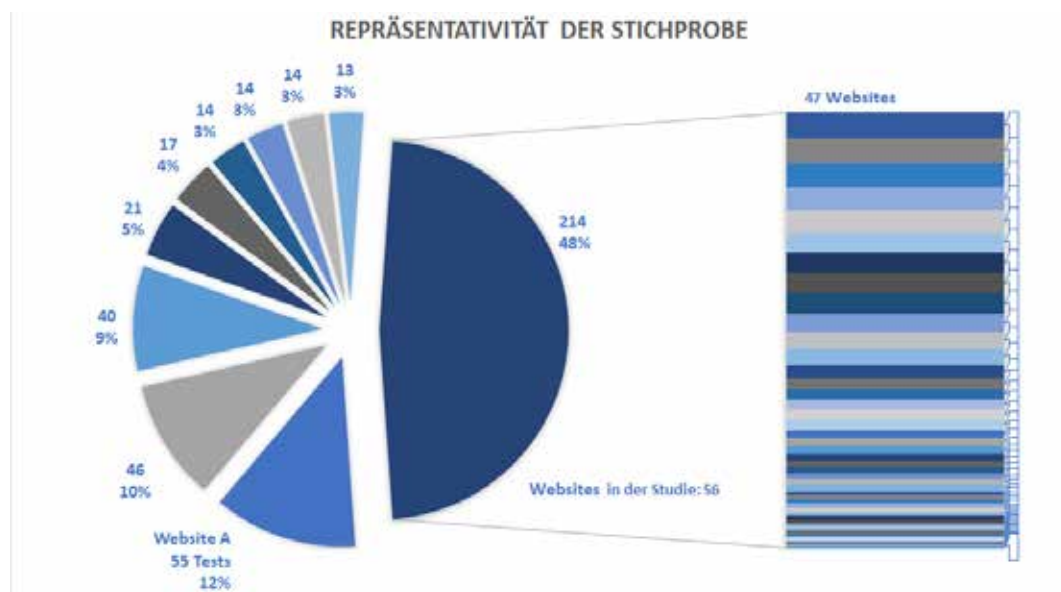


Abbildung 8 Verteilung der Stichprobe auf einzelne Websites

## 4.2 Welche Faktoren innerhalb des CRO-Prozesses beeinflussen den Erfolg?

Dieses Kapitel präsentiert die Studienergebnisse, die zur Beantwortung der ersten Forschungsfrage (vgl. S.2 und S.10) dienen. Dazu werden zunächst all diejenigen Faktoren präsentiert, die eine statistische Abhängigkeit in Bezug zur Erfolgswahrscheinlichkeit und/oder zu dem erzielten Uplift der Split-Tests aufgewiesen haben.

Die statistische Überprüfung auf Unabhängigkeit zwischen der Erfolgswahrscheinlichkeit und den jeweiligen Merkmalen erfolgte mithilfe des Chi2-Unabhängigkeitstests. Die statistische Signifikanz der unterschiedlichen Varianzen im erzielten Uplift wurde mit einem Zwei-Stichproben-t-Test unter der Annahme gleicher Varianzen berechnet.

Die Ergebnisse werden in Tabellenform präsentiert. Die Tabellen sind dabei inhaltlich in zwei Bereiche geteilt. Die Kennzahlen der linken Hälfte beschreiben den Zusammenhang der einzelnen Faktoren zur Effektivität, also der Wahrscheinlichkeit, ein Testergebnis mit positivem Uplift zu erhalten. In der rechten Seite der Tabellen wird der Zusammenhang mit der Effizienz dieser Faktoren beschrieben, also der Frage nachgegangen, wie hoch die durchschnittlich erzielten Uplifts bei Tests einer bestimmten Merkmalsklasse waren und wie die Unterschiede statistisch zu bewerten sind.

### Kontext- und Umfeldfaktoren in der CRO

In der Stichprobe befanden sich fast ausschließlich Websites aus den E-Business-Geschäftsmodellen „Commerce“ und „Connection“. In Tabelle 2 werden die wichtigsten Kennzahlen zu den einzelnen Merkmalen übersichtlich dargestellt.

336 der untersuchten Tests stammen von Websites des „Commerce“-Geschäftsmodells. Das bedeutet, dass die primäre Conversion dieser Websites in der Regel mit einer Transaktion zum Zweck eines Warenkaufs beschrieben werden kann. Bei den 108 Tests aus dem Geschäftsmodell „Connection“ hingegen lässt sich in der Regel bereits ein Weiterklicken zu einem Affiliate-Partner oder das Ausfüllen und Absenden eines Formulars als primäres Ziel der Website beschreiben.

Wie in Tabelle 2 zu sehen ist, haben die Tests der Klasse „Connection“ eine leicht höhere Erfolgsrate von 46,3%, im Vergleich zu 39,9% der Klasse „Commerce“. Das Signifikanzniveau liegt bei  $p = 0,20$ , weswegen keine direkte Abhängigkeit nachgewiesen werden konnte. Die Verteilung der erzielten Uplifts unterscheidet sich jedoch statistisch höchst signifikant ( $p < 0,01$ ).

## Geschäftsmodell

### E-Connection > 100% höhere Uplift als im E-Commerce

Während die Tests auf „Commerce“-Websites einen mittleren Uplift in Höhe von 12,6% (Standardabweichung 13,7) erreichen konnten, fällt der mittlere Uplift von 25,5% und einer Standardabweichung von 31 bei Tests der Klasse „Connection“ mehr als doppelt so hoch aus.

Die Daten zu den unterschiedlichen Geschäftsmodellen konnten die Nullhypothese, dass es also keinen statistischen Unterschied der Erfolgswahrscheinlichkeit unter Abhängigkeit des zugrunde liegenden Geschäftsmodells gibt, nicht widerlegen. Die Daten unterstützen jedoch die zweite Alternativhypothese, dass es einen höchst signifikanten Unterschied im durchschnittlich erzielten Uplift zwischen den beiden Stichproben gibt.

Geschäftsmodell	Erfolg					Uplift			
	Gesamt	Erfolg	Misserfolg	Erfolgsrate	Chi 2 - Unabhängigkeit p-Wert	Mittelwert	Median	Standardabweichung	P{T<=t} zweiseitig
Commerce	336	134	202	39,9%	0,20	12,6	8,0	13,7	0,00
Connection	108	50	58	46,3%		25,5	15,2	31,0	0,00
Sonstige	4	3	1	75,0%	x	x	x	x	x
<b>Gesamt</b>	<b>448</b>	<b>187</b>	<b>261</b>	<b>41,7%</b>		<b>16,1</b>	<b>9,5</b>	<b>20,6</b>	

Tabelle 2 Statistische Abhängigkeiten zu den vorhandenen Geschäftsmodellen



Branche	Erfolg					Uplift			
	Gesamt	Erfolg	Misserfolg	Erfolgsrate	Chi 2 - Unabhängigkeit p-Wert	Mittelwert	Median	Standardab- weichung	P(T<=t) zweiseitig
Automobile	46	23	23	50,0%	0,23	24,2	20,6	23,7	0,08
Dienstleistungen	14	5	9	35,7%	0,64	10,0	x	x	x
Konsumgüter	82	35	47	42,7%	0,85	16,9	12,3	16,1	0,83
Finanzen & Versicherungen	73	29	44	39,7%	0,70	23,8	11,5	35,0	0,08
Handwerk, Haushalt & Garten	61	22	39	36,1%	0,33	10,5	4,8	14,7	0,22
Industriegüter	24	8	16	33,3%	0,39	19,2	x	x	x
Kleidung & Accessoires	85	39	46	45,9%	0,39	8,3	5,5	9,8	0,02
Urlaub & Reisen	25	10	15	40,0%	0,9	17,5	11,9	18,3	0,83
Sonstige	38	16	22	42,1%	1,0	11,9	x	x	x
<b>Gesamt</b>	<b>448</b>	<b>187</b>	<b>261</b>	<b>41,7%</b>		<b>16,1</b>	<b>9,5</b>	<b>20,6</b>	

Tabelle 3 Abhängigkeiten nach untersuchten Branchen

Die in der Studie untersuchten Branchen konnten keine signifikanten Abhängigkeiten in Bezug auf die Erfolgswahrscheinlichkeit belegen. Bei der Betrachtung der im Mittel erzielten Uplifts nach Branchen zeigt sich ein signifikanter Downcast bei Websites der Branche „Kleidung & Accessoires“. Ein mittlerer Uplift in dieser Branche von 8,3% bei einer Standardabweichung von 9,8 bedeutet im Schnitt beinahe 50% weniger als der durchschnittliche Uplift aller Tests. Auffällig sind zusätzlich die positiven Tendenzen zu höheren Uplifts in den Branchen „Automobile“ und „Finanzen & Versicherungen“, welche bei einem Signifikanzniveau von jeweils 0,08 noch nicht ausreichend belastbar sind.

Die Untersuchung der Umfeldfaktoren „Geschäftsmodell“ und „Branche“ konnte zumindest vereinzelte Merkmale ausfindig machen, die laut den zur Verfügung stehenden Daten in einem statistisch signifikanten Zusammenhang mit den Erfolgsdeterminanten eines CRO-Prozesses stehen. Insbesondere sind dabei die höchst signifikanten Unterschiede zwischen den erzielten Uplifts nach unterschiedlichen Geschäftsmodellen und bei Tests aus der Branche „Kleidung & Accessoires“ zu nennen.



### Faktoren der Analyse-Phase

Nach den Umfeldfaktoren werden in den folgenden Abschnitten die Faktoren innerhalb des CRO-Prozesses vorgestellt, die laut der Studie im Zusammenhang mit der Effektivität und Effizienz dieser Prozesse stehen.

Zunächst werden die Daten des Faktors „Hypothesenursprung“ betrachtet. Da die Dokumentation der untersuchten Split-Tests in den meisten Fällen nicht ausreichte, um den Ursprung der Hypothese zu bestimmen, konnte dieses Merkmal für den Großteil aller Tests nicht erfasst und analysiert werden.

Einzig der Hypothesenursprung „Tweak“, also die Weiterentwicklung eines zuvor bereits getesteten Konzepts, konnte für alle Tests dieser Kategorie bestimmt werden. Tabelle 4 zeigt einen Vergleich der beiden Stichproben von Tests, die von einem „Tweak“ stammen, und allen anderen „Nicht-Tweaks“ (! Tweak). Die Daten zeigen eine leichte Tendenz dazu, dass die Erfolgswahrscheinlichkeit von Tweaks niedriger sein könnte. Der Unterschied in den Erfolgsraten ist jedoch nicht ausreichend, um bei dem vorhandenen Stichprobenumfang signifikante Unterschiede zu belegen.

Als weiterer Faktor der Analyse-Phase eines CRO-Prozesses wurden die einzelnen Seitentypen untersucht, auf denen die Optimierung stattfand. Testkonzepte, welche auf mehr als einem Seitentyp umgesetzt wurden, werden in der Analyse nicht bewertet, da jede Kombinationsmöglichkeit als eigenständige Merkmalsausprägung hätte gewertet werden müssen. In der Stichprobe wären diese einzelnen Ausprägungen zu selten vorgekommen, um diese einer statistischen Bewertung zu unterziehen. Wenn es sich jedoch um Änderungen handelte, die auf allen oder annähernd allen Seiten der testenden Website umgesetzt wurden, erfolgte eine Zuordnung zur Seitentyp-Klasse „Global“. In Tabelle 5 finden sich alle möglichen Merkmalsklassen des Faktors „Seitentyp“ und die jeweiligen statistischen Leistungswerte.

Hypothesen- Ursprung	Erfolg					Uplift			
	Gesamt	Erfolg	Misserfolg	Erfolgsrate	Chi 2 - Unabhängigkeit p-Wert	Mittelwert	Median	Standardab- weichung	P(T<=t) zweiseitig
Tweak	46	15	31	32,6%	0,20	17,7	10,1	22,2	0,77
!Tweak	400	170	230	42,5%		16,1	9,5	20,6	0,77
<b>Gesamt</b>	<b>448</b>	<b>187</b>	<b>261</b>	<b>41,7%</b>		<b>16,1</b>	<b>9,5</b>	<b>20,6</b>	

Tabelle 4 Zusammenhang zwischen dem Hypothesenursprung und der Erfolgswahrscheinlichkeit sowie Effizienz



Seitentyp	Erfolg					Uplift			
	Gesamt	Erfolg	Misserfolg	Erfolgsrate	Chi 2 - Unabhängigkeit p-Wert	Mittelwert	Median	Standardab- weichung	P(T<=t) zweiseitig
Startseite	18	7	11	38,9%	0,21	19,8	x	x	x
Landingpages	35	14	21	40,0%	0,83	17,4	11,8	11,6	0,82
Kategorieseiten	53	20	33	37,7%	0,08	18,4	13,2	29,2	0,65
Ergebnisseiten	50	25	25	50,0%	0,21	18,7	10,1	19,8	0,55
PDS	100	42	58	42,0%	0,95	10,9	7,9	8,7	0,11
Konfiguratoren	55	29	26	52,7%	0,08	22,5	14,2	25,7	0,14
Warenkorb & Checkout	52	13	39	25,0%	0,01	19,6	6,8	21,3	0,55
Global	48	21	27	43,8%	0,77	7,4	6,1	6,5	0,06
<b>Gesamt</b>	<b>448</b>	<b>187</b>	<b>261</b>	<b>41,7%</b>		<b>16,1</b>	<b>9,5</b>	<b>20,6</b>	

Tabelle 5 Seitentypen und ihr Einfluss auf die Erfolgsdeterminanten

Es zeigte sich, dass Tests auf dem Seitentyp „Warenkorb/Checkout“ eine höchst signifikant schlechtere Erfolgsrate verzeichnen als auf allen anderen Seitentypen. Die Daten aus Tabelle 5 weisen zusätzlich sowohl auf eine leicht negative Tendenz in der Erfolgswahrscheinlichkeit auf Kategorie-Seiten als auch auf eine positive Performance von Konfiguratoren hin. Beide Werte sind bei einem Signifikanzniveau von 0,08 jedoch nicht vollständig belastbar.

Die positive Tendenz von Konfigurator-Seiten zeigt sich ebenfalls bei der Betrachtung der erreichten Uplifts. Ein mittlerer Uplift von 22,5% auf Konfigurator-Seiten bedeutet eine deutliche Steigerung zum Gesamtschnitt von 16,1%, auch wenn die Ergebnisse nicht signifikant sind.

Auffällig sind außerdem insbesondere die deutlich schlechteren Leistungsdaten der Seitentyp-Klasse „Global“. Ein mittlerer Wert von 7,4% bei einer geringen Standardabweichung von 6,5 bedeutet die schlechteste Effizienz aller untersuchten Seitentypen, bei einem annähernd signifikanten Ergebnis (Signifikanzniveau 0,06).

Bei der Bewertung dieser Ergebnisse müssen jedoch weitere mögliche Abhängigkeiten und Muster berücksichtigt werden. So ist es bspw. offensichtlich, dass der Seitentyp Warenkorb/Checkout, in der Regel nur Websites vorhanden ist, die der Geschäftsmodell-Klasse Commerce angehören. Eine segmentierte Betrachtung nach unterschiedlichen Zieltypen (Micro- und Macro Conversions) kann ein weiterer Ansatzpunkt für eine genauere Betrachtung des beobachteten Phänomens sein, da auf dem Seitentyp Warenkorb/Checkout in der Regel keine Micro-Conversions als Ziel bestimmt werden.

Intention	Erfolg					Uplift			
	Gesamt	Erfolg	Misserfolg	Erfolgsrate	Chi 2 - Unabhängigkeit p-Wert	Mittelwert	Median	Standardab- weichung	P(T<=t) zweiseitig
Motivation (M)	79	30	49	38,0%	0,21	10,5	8,3	7,8	0,14
Ability (A)	30	5	25	16,7%	0,00	8,0	x	x	x
Trigger (T)	113	45	68	39,8%	0,63	13,5	6,5	18,9	0,44
MA	6	3	3	50,0%	x	12,5	x	x	x
MT	105	54	51	51,4%	0,02	18,5	11,3	22,8	0,47
MAT	26	13	13	50,0%	0,38	24,3	18,6	22,9	0,18
AT	89	37	52	41,6%	0,48	18,9	11,6	25,1	0,47
Gesamt	448	187	261	41,7%		16,1	9,5	20,6	

Tabelle 6 Intentionen der Hypothese und deren Leistungsdaten

### Faktoren der Hypothesen-Phase

Tabelle 6 präsentiert die Ergebnisse des Faktors „Intention“. Die Klassen definieren sich aus dem Behavior-Modell nach Fogg und beschreiben, wie die getesteten Änderungen der Hypothese auf das Verhalten der Nutzer wirken und dadurch zu einer Konversion führen sollen.

In Bezug auf die Erfolgswahrscheinlichkeit lieferte die Studie zwei signifikante Ergebnisse. Zum einen haben Tests, deren Intention einzig in der Verbesserung der Ability liegt, eine um mehr als 60% niedrigere Erfolgsrate als der Durchschnitt (16,67% zu 41,74%). Zum anderen ist die Erfolgswahrscheinlichkeit bei einer Kombination der Intentionen Motivation und Trigger (MT) mit 51,43% merklich höher als im Mittel.

Die Daten aus Tabelle 6 zeigen außerdem eine Tendenz dazu, dass alle Kombinationsmöglichkeiten, welche Motivation mit einem weiteren Merkmal verbinden, eine höhere Erfolgsrate zu haben scheinen. Unterschiede in der Effizienz der einzelnen Intention, also welche Intentionen

im Schnitt höhere Uplifts erzielen, konnten in den untersuchten Daten nicht gefunden werden.

Nicht nur die Intention einer Hypothese zeigte eine Abhängigkeit zur Erfolgswahrscheinlichkeit. Auch die dazu verwendeten Stimuli waren Teil der Untersuchung. Da die Stichprobe eine große Anzahl unterschiedlicher Ausprägungen dieses Merkmals enthielt, konnten keine sinnvollen Untersuchungen bezüglich spezieller Stimuli durchgeführt werden. Jedoch konnte ein Zusammenhang hergestellt werden zwischen der Anzahl verwendeter Stimuli und der Erfolgswahrscheinlichkeit.

Wie Abbildung 9 zeigt, scheint es einen direkten linearen Zusammenhang zwischen der Anzahl der Stimuli und der Erfolgsrate zu geben. Der hohe Unterschied in der Erfolgswahrscheinlichkeit bei zwei verwendeten Stimuli (Spalte „n = 2“ im Diagramm) erweist sich nach Überprüfung mithilfe eines Chi2-Unabhängigkeitstests als höchst signifikant ( $p = 0,01$ ). Für eine Überprüfung der größeren Klassen reichten die einzelnen Stichproben nicht mehr aus.

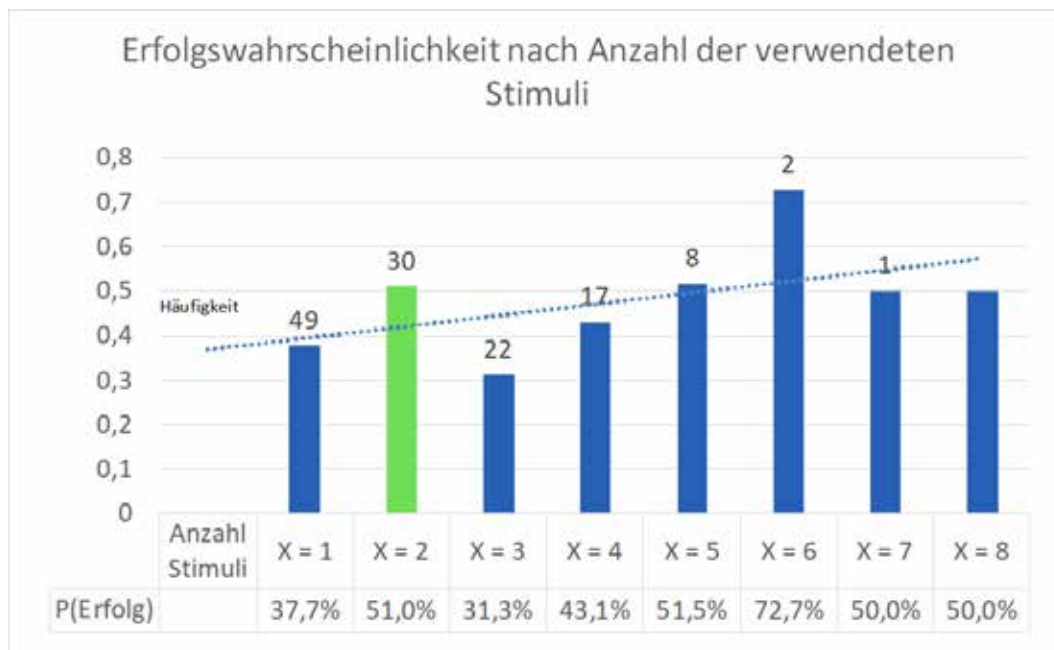


Abbildung 9 Abhängigkeit der Erfolgswahrscheinlichkeit von der Anzahl verwendeter Stimuli

Bei der Betrachtung der unterschiedlichen Zieltypen (vgl. Tabelle 7) zeigt sich eine höchst sig-nifikant höhere Erfolgsrate bei Tests, deren Ziel das Erreichen eines sekundären Ziels, also einer Micro Conversion, war. Tests der Zieltyp-Klasse „Micro“ hatten eine Erfolgsrate von 64,7%, während „Macro“-Ziele nur in 38,9% aller Fälle ein positives Testergebnis hatten.

Die interessante Frage nach dem Unterschied zwischen qualitativen und quantitativen Macro-Zielen konnte nicht untersucht werden, da die vorhandene Dokumentation eine Betrachtung der qualitativen Ziele (bspw. Revenue/Visitor) nicht zuließ.

Ziel-Typ	Erfolg					Uplift			
	Gesamt	Erfolg	Misserfolg	Erfolgsrate	Chi 2 - Unabhängigkeit p-Wert	Mittelwert	Median	Standardabweichung	P(T<=t) zweiseitig
Macro	396	154	242	38,9%	0,00	26,0	8,8	93,4	0,98
Micro	51	33	18	64,7%		26,4	20,6	30,5	
Gesamt	448	187	261	41,7%		16,1	9,5	20,6	

Tabelle 7 Unterschiede der Split-Tests nach Zieltypen

### Weitere Faktoren

Zum Abschluss der Betrachtung der einzelnen Faktoren innerhalb des CRO-Prozesses werden im folgenden Abschnitt die übrigen Merkmale präsentiert, die statistische Auffälligkeiten in der Analyse aufgewiesen haben.

Zunächst werden die Daten des Faktors „Hypothesenursprung“ betrachtet. Da die Dokumentation der untersuchten Split-Tests in den meisten Fällen nicht ausreichte, um den Ursprung der Hypothese zu bestimmen, konnte dieses Merkmal für den Großteil aller Tests nicht erfasst und analysiert werden.

Bei der Interpretation dieser Ergebnisse muss jedoch beachtet werden, dass die Anzahl der getesteten Varianten keiner konzeptuellen Entscheidung im CRO-Prozess zugrunde liegen muss, da die Anzahl der möglichen Varianten im Test auch durch den zur Verfügung stehen-den Traffic technisch beschränkt ist.

Deswegen können die Ergebnisse Ausdruck eines direkten kausalen Zusammenhangs sein, der sich etwa dadurch erklären ließe, dass das Potenzial einer Hypothese stärker genutzt wird, wenn sich daraus zwei oder mehr Varianten ableiten lassen.

Andererseits kann es sich bei den beobachteten Abhängigkeiten um eine Korrelation handeln, die sich beispielsweise damit erklären ließe, dass Websites mit hohem Traffic und der daraus resultierenden Möglichkeit, mehr Varianten zu testen, eine höhere Erfolgswahrscheinlichkeit besitzen. Außerdem gilt aus statistischer Sicht zu beachten, dass bei Experimenten mit mehr Varianten die Wahrscheinlichkeit eines Alpha-Fehlers steigt, also dem Fall, dass eine Variante im Split-Test ein signifikant positives Ergebnis erzielt, obwohl die Variante in Wirklichkeit schlechtere Ergebnisse liefert. (vgl. Sharma, 2015, S. 297)

Anzahl (n) Varianten	Erfolg					Uplift			
	Gesamt	Erfolg	Misserfolg	Erfolgsrate	Chi 2 - Unabhängigkeit p-Wert	Mittelwert	Median	Standardab- weichung	P(T<=t) zweiseitig
n=1	264	98	166	37,1%		17,5	10,1	22,3	
n>1	182	87	95	47,8%	0,02	36,2	8,3	122,7	0,07
Gesamt	448	187	261	41,7%		16,1	9,5	20,6	

Tabelle 8 Anzahl der getesteten Varianten und deren Auswirkung auf die Erfolgskennzahlen

Als letzten Faktor des CRO-Prozesses, der relevante Ergebnisse in der Studie liefern konnte, werden in Tabelle 9 die unterschiedlichen Verteilungen nach dem Faktor Devices präsentiert. Die Daten zeigen für Tests auf mobilen Seiten und Apps eine klare Tendenz zu einer erhöhten Erfolgswahrscheinlichkeit (Signifikanzniveau 0,08) und einen signifikant höheren mittleren Uplift. Der im Schnitt deutlich niedrigere Uplift von reinen „Desktop“-Tests erweist sich sogar als höchst signifikant.

Wie bereits die zuvor untersuchten Umfeldfaktoren, konnten auch innerhalb des CRO-Prozesses einzelne Merkmale sowie spezielle Ausprägungen dieser Faktoren ausgemacht werden, welche in der Studie signifikante Unterschiede in der Erfolgswahrscheinlichkeit (Effektivität) und den im Durchschnitt erzielten Uplifts (Effizienz) aufweisen.

Device	Erfolg					Uplift			
	Gesamt	Erfolg	Misserfolg	Erfolgsrate	Chi 2 - Unabhängigkeit p-Wert	Mittelwert	Median	Standardab- weichung	P(T<=t) zweiseitig
Desktop	370	153	217	41,4%	0,72	14,9	8,1	26,4	0,00
Mobile	32	18	14	56,3%	0,08	64,1	13,1	196,3	0,05
Responsive	46	16	30	34,8%	0,31	90,3	32,0	170,2	0,63
Gesamt	448	187	261	41,7%		16,1	9,5	20,6	

Tabelle 9 Abhängigkeiten der Tests für Mobile und/oder Desktop Traffic

### 4.3 Beziehungen und Muster zwischen den Faktoren selbst

In diesem Teil der Auswertung werden die Ergebnisse präsentiert, anhand derer die Forschungsfragen 2 und 3 (vgl. S.2 und S.10) beantwortet werden sollen. Ziel der Analysen war es, bestimmte Konstellationen innerhalb der untersuchten Variablen zu finden, welche Auswirkungen auf die Erfolgswahrscheinlichkeit haben.

Um die Frage nach den Beziehungen unter den Faktoren selbst zu beantworten, wurden mathematische Cluster für erfolgreiche und nicht-erfolgreiche Tests, mithilfe des k-Means-Clustering-Algorithmus gebildet. Als mögliches Vorhersagemodell wird anschließend ein logistisches Regressionsmodell vorgestellt, welches die Wahrscheinlichkeit eines erfolgreichen Tests anhand mehrerer Faktoren bzw. Prädiktoren berechnet.

#### 4.3.1 Welche Hypothesen sind erfolgsversprechend

Zunächst wurde ein Schwerpunkt auf eine genauere Betrachtung der Zusammenhänge zwischen den Hypothesenfaktoren „Intention“ und „Stimulus/i“ gelegt, um bestimmte Konstellationen auszumachen, welche die Erfolgswahrscheinlichkeit relevant beeinflussen. Dazu wurde die Stichprobe in zwei Klassen geteilt, für erfolgreiche und nicht erfolgreiche Tests, und diese beiden Klassen jeweils einer k-Means-Clustering-Analyse unterzogen.

Es zeigte sich, dass aufgrund der relativ vielen Ausprägungen (12 verschiedene Stimuli wurden erfasst) des Stimulus-Faktors der Stichprobenumfang (NErfolg = 187) nicht ausreichte, um belastbare Ergebnisse zu erlangen. Dennoch lassen sich aus den erstellten Clustern Tendenzen für bestimmte Hypothesen erkennen. Tabelle 10 zeigt die Ergebnisse der durchgeführten k-Means-Clustering-Analyse. Insgesamt wurden 187 erfolgreiche Tests in sechs Cluster (k = 6) und 261 nicht erfolgreiche Tests in sieben Cluster (k = 7) aufgeteilt. Eine ausführliche Auswertung inklusive aller absoluten Häufigkeiten der einzelnen Prädiktoren befindet sich im Anhang dieser Arbeit.

	k-Means Cluster für erfolgreiche Tests							k-Means Cluster für nicht erfolgreiche Tests						
	Cluster e0	Cluster e1	Cluster e2	Cluster e3	Cluster e4	Cluster e5		Cluster n0	Cluster n1	Cluster n2	Cluster n3	Cluster n4	Cluster n5	Cluster n6
Größe (N)	55	37	39	3	21	32		87	17	26	31	24	9	67
Intention	T	M	MT	AT	MT	AT		MT	T	M	T	M	T	AT
Banner	no	no	no	no	no	no		no	no	no	no	no	no	no
Bewertungen	no	no	no	no	no	no		no	no	no	no	no	no	no
Design	no	no	yes	no	no	yes		no	no	yes	no	yes	no	yes
Icons	no	no	no	no	no	no		no	no	no	no	no	no	no
Layout	no	no	yes	no	no	yes		no	no	yes	no	yes	no	yes
Navigation	no	no	no	no	no	no		no	no	no	no	no	no	no
Nudges	yes	no	no	no	no	no		no	yes	no	no	yes	no	no
PopUps/Layer	no	no	no	no	no	no		no	no	no	no	no	no	no
Produktsortierung	no	no	no	yes	no	no		no	no	no	no	no	no	no
Seitenstruktur	no	no	yes	yes	no	yes		no	no	yes	no	yes	no	yes
UVPs	no	no	yes	no	no	no		no	no	yes	no	no	yes	no
Wording	no	no	yes	no	no	no		no	no	yes	yes	no	no	no

Tabelle 10 Clusterbildung unter den Hypothesenfaktoren

Unter den berechneten Clustern für erfolgreiche und nicht erfolgreiche Tests wurden jeweils zwei Cluster ausgemacht, von denen sich mögliche Hypothesen zur Beantwortung der Forschungsfragen ableiten lassen.

Die einzelnen Cluster werden in Tabelle 10 spaltenweise abgebildet. Die linke Hälfte der Tabelle zeigt die sechs Cluster für erfolgreiche Tests und auf der rechten Hälfte der Tabelle sind die Cluster der nicht erfolgreichen Tests zu sehen.

Cluster e0 unter den erfolgreichen Tests beschreibt eine statistische Nähe zwischen der Hypothesen-Intention Trigger und dem gewählten Stimulus Nudges. Cluster e2 legt Rückschlüsse auf einen Zusammenhang zwischen der kombinierten Intention Motivation/Trigger sowie den Stimuli Design, Layout, Seitenstruktur, UVPs und Wording nahe.

Jedoch fällt auf, dass das Cluster n2 dieselben Stimuli in Kombination mit der Intention Motivation zusammenfasst und in statistischer Nähe zu nicht erfolgreichen Tests steht. Dies könnte ein Hinweis darauf sein, wie bedeutend die Kombination der Intentionen Motivation und Trigger für die Effektivität der Stimuli Design, Layout, Seitenstruktur, UVPs und Wording ist.

Cluster n4 beschreibt eine statistische Nähe zwischen der Intention Motivation und den Stimuli Design, Nudges, Layout und Seitenstruktur.

Wie beschrieben und in Tabelle 11 zu sehen, gibt es unter den gebildeten Clustern noch zu viele Überschneidungen, die eine detailliertere Auswertung verhindern. Durch eine größere Stichprobe und/oder eine Anpassung der untersuchten Variablen kann das Modell zukünftig weiter verfeinert werden, um belastbare und aussagekräftigere Informationen daraus zu erhalten.

Auf Basis der Studie ließen sich jedoch bereits erkennen, dass eine Clusterbildung unter den ausgewählten Faktoren grundsätzlich zur Bestimmung von erfolgsbestimmenden Zusammenhängen geeignet sein kann.

### 1.1.1 Modell zur Vorhersage der Erfolgswahrscheinlichkeit

Um die Faktoren des Modells daraufhin zu überprüfen, ob diese sich zur Vorhersage der Erfolgswahrscheinlichkeit eignen, wurde eine logistische Regressionsanalyse durchgeführt.

Für die Untersuchung eigneten sich folgende Faktoren, im Sinne der Regression als Prädiktoren bezeichnet: Tweak, n-Stimuli, Geschäftsmodell und Seitentyp. Als Kriterium der Untersuchung wird der Erfolg eines Experiments als binominale abhängige Variable gewählt. Tabelle 11 zeigt die Koeffizienten der einzelnen Prädiktoren und die zugehörige Statistik.

Logit	Variable	Coeff.	Std. Err.	z-score	P> z
yes	Tweak	-0,592	0,387	-1,529	0,13
yes	Mittel Anzahl	0,088	0,074	1,188	0,23
yes	Commerce=1	-1,570	1,160	-1,353	0,18
yes	Connection=1	-1,352	1,180	-1,146	0,25
yes	Startseite	0,105	0,440	0,238	0,81
yes	Landingpages	-0,252	0,359	-0,700	0,48
yes	Kategorieseiten	0,363	0,289	1,255	0,21
yes	Ergebnisseite / Listing	-0,074	0,297	-0,249	0,80
yes	PDS	-0,012	0,250	-0,049	0,96
yes	Konfiguratoren	0,326	0,307	1,063	0,29
yes	Checkout/Warenkorb	-0,531	0,316	-1,683	0,09
yes	Constant	1,011	1,185	0,854	0,39

Tabelle 11 Statistik und Koeffizienten zur logistischen Regression

Die p-Werte in der Tabelle zeigen, dass das Modell noch deutlichen Zufallseffekten zugrunde liegen kann. Die Trainingsdaten für das Modell bestanden aus 80% (397 Experimente) der gesamten Stichprobe. Die übrigen 20% aller Tests dienten zur Bewertung des Regressionsmodells als Testdaten. Anhand der Testdaten wurden die Vorhersagen des erzeugten logistischen Regressionsmodells überprüft und mit den tatsächlichen Ergebnissen verglichen. Die Ergebnisse dieser Untersuchung werden in der Klassifikationsmatrix in Tabelle 12 präsentiert.

Beobachtetes Ergebnis		Vorhergesagtes Ergebnis		korrekt vorhergesagt
		Erfolg		
		ja	nein	
Erfolg	ja	8	29	21,62%
	nein	7	46	86,79%
Summe				60%

Tabelle 12 Klassifikationsmatrix für die logistische Regression



Wie in Tabelle 12 zu sehen ist, konnte das Ergebnis von insgesamt 60% aller Testdaten durch das Modell korrekt vorhergesagt werden. Um das Ziel einer Effektivitätssteigerung von CRO-Prozessen zu erreichen, ist natürlich insbesondere die Vorhersage erfolgreicher Tests relevant. Da von 37 erfolgreichen Testdaten acht vom Modell korrekt vorhergesagt wurden, ist die Genauigkeit für diesen Anwendungsfall mit 21,62% zu niedrig, um sinnvoll genutzt zu werden.

Auch wenn die Datengrundlage kein belastbares Modell zur Vorhersage der Erfolgswahrscheinlichkeit in Abhängigkeit mehrerer Einflussfaktoren ergeben hat, beschreibt die Studie in dieser Richtung dennoch das Potenzial einer logistischen Regression auf Basis einiger Faktoren des Untersuchungsmodells.

## **4.4 Beantwortung der Forschungsfragen**

In diesem Kapitel werden die Ergebnisse der Auswertung zusammengefasst, um die konkreten Forschungsfragen der Arbeit zu beantworten.

### **4.4.1 Forschungsfrage 1**

Gibt es kategorisierbare Merkmale des CRO-Prozesses, zu denen die Erfolgswahrscheinlichkeit und der zu erwartende Uplift der durchgeführten Split-Tests in statistischer Abhängigkeit stehen, und wenn ja, welche?

Ja, die Analyse konnte mehrere Faktoren innerhalb des CRO-Prozesses und aus dessen Umfeld ausmachen, die in statistischer Abhängigkeit zur Effektivität und Effizienz dieser Prozesse zu sein scheinen.

Die Auswertung hat ergeben, dass Tests auf den Seitentypen Warenkorb/Checkout sowie Hypothesen mit der Intention, den Ability-Faktor zu verbessern, eine höchst signifikant niedrigere Erfolgsquote haben. Wohingegen Hypothesen, die zur Intention eine Kombination aus Motivation und Trigger (im Sinne des Behavior-Modells) haben, eine signifikant höhere Erfolgswahrscheinlichkeit aufweisen.

Verfolgt eine Hypothese den Zweck, auf ein sekundäres Ziel (Micro Conversions) hin zu optimieren, ist die Erfolgswahrscheinlichkeit signifikant höher als bei einer Optimierung auf primäre Conversion-Ziele.

Optimierungen auf mobilen Websites zeigen sowohl positive Tendenzen in der Erfolgswahrscheinlichkeit als auch einen signifikanten Unterschied bei der Effizienz unter diesen erfolgreichen Tests.

Aus dem Umfeldfaktor „Branche“ konnte insbesondere die Branche „Kleidung & Accessoires“ ausgemacht werden, die in der Stichprobe im Schnitt signifikant niedrigere Uplifts erzeugt als im gesamten Durchschnitt.

Generell scheinen E-Commerce-Websites niedrigere Uplifts zu erzeugen als Tests von Websites des Geschäftsmodells „Connection“.

Alle Aussagen über die Zusammenhänge der einzelnen Faktoren sind als Ablehnung der Nullhypothese zu betrachten, welche besagt, dass die Kennzahlen der Effektivität/Effizienz in den verglichenen Stichproben derselben Grundgesamtheit entstammen und keine signifikanten Unterschiede bestehen.

#### 4.4.2 Forschungsfrage 2

Gibt es eine statistische Beziehung zwischen mehreren dieser Faktoren?

Diese Frage konnte nicht abschließend beantwortet werden, auf Basis einer k-Means-Clusteranalyse ergaben sich jedoch Hinweise auf sowohl erfolgsversprechende als auch erfolgshemmende Muster unter den beiden Faktoren Stimulus und Motivation. Die Ergebnisse geben Grund zu der Annahme, dass eine weitere Studie mit einer anderen Stichprobe die erkannten Trends bestätigen und weitere Zusammenhänge entdecken könnte.

#### 4.4.3 Forschungsfrage 3

Gibt es statistische Beziehungen zwischen mehreren dieser Faktoren, die eine Vorhersage der Erfolgswahrscheinlichkeit erlauben?

Auch diese Frage konnte in der Studie nicht eindeutig beantwortet werden. Die Studie zeigt jedoch die grundsätzliche Möglichkeit einer logistischen Regression mit bestimmten Faktorausprägungen des CRO-Prozesses. Durch eine größere Menge an Trainings- und Testdaten könnte das Modell ein zufriedenstellendes Gütemaß erreichen und eine Vorhersage der Erfolgswahrscheinlichkeit ermöglichen.

## 4.5 Kritische Bewertung der Forschungsergebnisse

Grundsätzlich gilt es bei der Bewertung aller Ergebnisse der Studie zu berücksichtigen, dass die Datengrundlage auf einer einzelnen Quelle beruht. Alle untersuchten Split-Tests wurden von der „Web Arts AG“ entwickelt oder in der Entwicklung zumindest betreut. Firmeneigene Prozesse, Methoden, Ideen und Erfahrungen der „Web Arts AG“ können daher Einfluss auf die Ergebnisse gehabt haben und deren Effekte sind ohne weitere Untersuchungen nicht zu bestimmen.

Die Auswahl der Stichprobe genügt keinen repräsentativen Kriterien. Wie in der Beschreibung der Stichprobe erwähnt, entstammen viele Tests der Stichprobe einigen wenigen Websites. Zudem konnten einzelne Experimente nicht in der Studie berücksichtigt werden, weil die vorhandene Dokumentation nicht ausreichte, um die Tests im Sinne des Untersuchungsmodells zu kategorisieren.

Unter Berücksichtigung dieser genannten Gründe, der beschriebenen Stichproben-Zusammensetzung aus Abschnitt 4.1, bei der die Hälfte aller untersuchten Split-Tests von neun Websites stammen, und dem Umstand des relativ kleinen Stichprobenumfangs, muss die Validität aller statistischen Ergebnisse dieser Arbeit mit weiteren Studien zunächst bestätigt werden.

Aus unternehmerischer Sicht sollte bei der Bewertung der Ergebnisse beachtet werden, dass, obwohl in der gesamten Arbeit die Optimierung der Conversion Rate als Geschäftsziel aufgefasst wurde, eine Fixierung auf diese Metrik auch Schwächen darin hat, Aussagen über den tatsächlichen unternehmerischen Erfolg zu tätigen. (Vgl. Sharma, 2015, S. 128)

Die Conversion Rate ist eine relative Kennzahl, welche die Häufigkeit von Transaktionen (Conversions) auf einer Website beschreibt. Eine Fixierung in der Optimierung auf diese Kennzahl kann zu Problemen in der Messung und Interpretation führen. Auch eine Verbesserung der Kennzahl ohne positive Auswirkung auf das Geschäftsergebnis wäre möglich. Eine deutliche Preissenkung von Produkten könnte die Conversion Rate einer E-Commerce-Seite etwa leicht erhöhen, ohne dabei auch den Deckungsbeitrag des Unternehmens zu steigern. (Vgl. Sharma, 2015, S. 141–143)

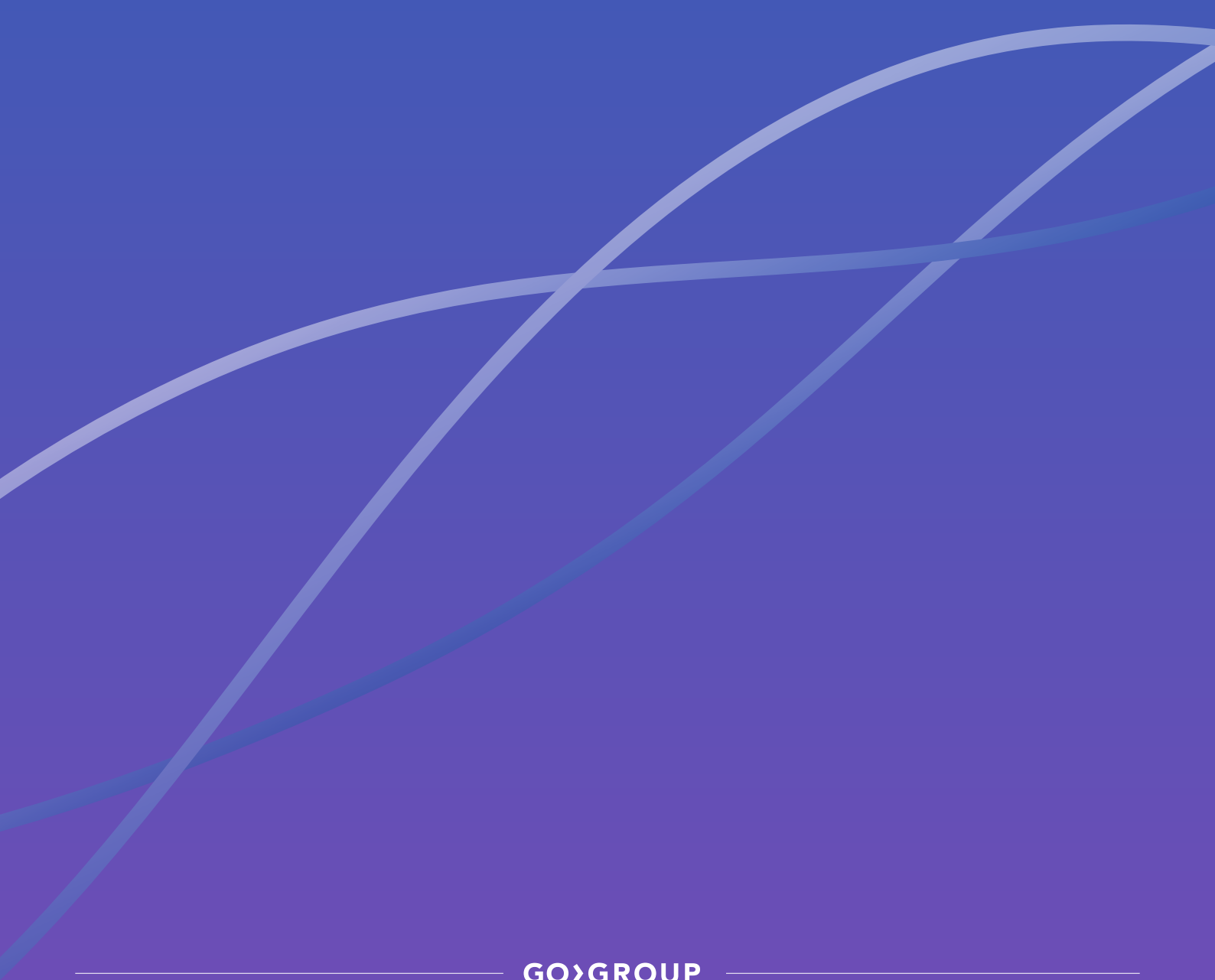
Um diese Geschäftsziele zu erreichen, werden in der Praxis häufig qualitative Ziele als Erfolgsparameter der Optimierung empfohlen, zum Beispiel der erwirtschaftete Gesamtumsatz, der durchschnittliche Umsatz pro Bestellung oder der positive Beitrag zum Geschäftsergebnis. (Vgl. Morys, 2011, S. 54 f.)

Wie bereits im Kapitel Forschungsmodell beschrieben, waren die notwendigen Daten zu diesen Erfolgskennzahlen häufig nicht vorhanden. Deswegen konnte diese Betrachtungsebene, trotz ihrer Relevanz, in dieser Arbeit nicht berücksichtigt werden.

Abschließend muss insbesondere bei der Bewertung der unterschiedlichen Uplifts darauf hingewiesen werden, dass in der Studie zur Berechnung ein t-Test verwendet wurde. Dazu wurde mit der Annahme gearbeitet, dass die Verteilung der in den einzelnen Tests erzielten Uplifts einer Gauß'schen Normalverteilung entsprechen. Die tatsächliche Verteilung kann in der Praxis jedoch deutlich abweichen. (Vgl. Sharma, 2015, S. 219)

Eine Alternative in diesem Fall wären sogenannte nonparametrische oder verteilungsfreie Tests, wie z. B. der Wilcoxon-Vorzeichen-Rang-Test. (vgl. Fahrmeir u. a., 2003, S. 436–441) Aus Zeitgründen musste in dieser Arbeit auf dieses Verfahren verzichtet und die Nachteile des weniger aufwendigen t-Test-Verfahrens in Kauf genommen werden.

# 05 FAZIT UND AUSBLICK





Das Ziel dieser Studie war es, ein Modell zu entwickeln, mit dem CRO-Prozesse im Unternehmen und deren Einflussfaktoren beschrieben und kategorisiert werden können, um damit den Einfluss dieser Faktoren auf die Effektivität und Effizienz in der CRO zu bestimmen. Dieses Ziel konnte erreicht werden und Hypothesen über bestimmte Abhängigkeiten innerhalb der CRO formulieren. Weitere Ziele waren einerseits, zu überprüfen, ob diese Faktoren dazu dienen können, bestimmte Voraussetzungen oder Muster zu erkennen, die eine hohe Erfolgsrate für Split-Tests begünstigen. Andererseits wollte diese Arbeit erörtern, ob sich anhand der untersuchten Faktoren statistische Vorhersagemodelle bilden lassen.

In diesem Fazit der Studie werden die wichtigsten Ergebnisse der Arbeit kurz zusammengefasst und zum einen erklärt, wie diese Forschungsergebnisse in der praktischen Anwendung nützlich sein könnten; zum anderen werden Ansätze hervorgehoben, die für weitere mögliche Forschungen relevant sein könnten.

Die Studienergebnisse zeigen signifikante Unterschiede in der Erfolgswahrscheinlichkeit je nach Seitentyp, auf welchem getestet wurde, der Anzahl der verwendeten Stimuli im Test oder der verhaltenspsychologischen Intention, welche die Hypothese verfolgt.

Bei der Betrachtung unterschiedlicher Seitentypen zeigte die Statistik zum einen, dass als Konfiguratoren klassifizierte Seitentypen eine besonders hohe Erfolgswahrscheinlichkeit haben. Dies könnte für Websites, deren Conversion Funnel solche Konfiguratoren beinhalten, ein relevantes Argument sein, einen Fokus bei der Optimierung auf genau diese Seitentypen zu legen.

Im Gegensatz zu den Konfiguratoren zeigen die Seitentypen Warenkorb/Checkout, die am Ende eines Conversion Funnel über mehrere Seiten stehen, eine deutlich schlechtere Erfolgswahrscheinlichkeit. Auf Basis dieser Ergebnisse könnte empfohlen werden, sich bei der Optimierung vor allem auf andere Seitentypen zu konzentrieren.

Ein Schwerpunkt der Studie lag auf der Betrachtung der im CRO-Prozess entwickelten Hypothesen. Die Intention der Hypothesen, also die Frage, wie eine Veränderung an der Website den Nutzer zur Konvertierungshandlung bewegen soll, wurde mit den Variablen Motivation, Ability und Trigger des Behavior-Modells nach Fogg klassifiziert und analysiert.

Die Ergebnisse zeigen zum einen, dass die Verwendung des Faktors Motivation signifikant höhere Erfolgswahrscheinlichkeiten als der Durchschnitt aller anderen Hypothesen erzielt, und zum anderen, dass die Verwendung des Behavior-Modellfaktors Ability eine deutlich geringere Erfolgswahrscheinlichkeit in Split-Tests verspricht. Verstärkt wird der positive Effekt der Motivation, wenn die Hypothese Motivation und weitere Faktoren des Behavior-Modells kombiniert.

Folgt man den Ergebnissen der Studie, ist diese ein wissenschaftlicher Hinweis darauf, dass in der CRO häufiger Erfolge durch Split-Tests erzielt werden können, indem die Motivation der Nutzer gesteigert wird, als etwa bei einer Optimierung der Usability.

Zusätzlich zu den abhängigen Erfolgswahrscheinlichkeiten konnte die Studie Schlüsse darüber ziehen, welche Merkmale des CRO-Prozesses Einfluss auf die im Durchschnitt zu erwartenden Uplifts eines Tests haben. So zeigte die Auswertung der Daten etwa, dass Split-Tests auf E-Commerce-Seiten einen signifikant niedrigeren Uplift erzielen. Das Gleiche trifft auf Websites der Branche „Kleidung & Accessoires“ zu.

Auch wenn dieser statistische Nachweis dem betreffenden Unternehmen keinen unmittelbaren Nutzen bringt, um die Höhe seiner Uplifts zu steigern, kann diese Information zumindest bei der Bewertung der eigenen Leistungsnachweise dienlich sein und soll vor allem ein Anhaltspunkt für weitere Untersuchungen sein, um die genaueren Ursachen dieser Zusammenhänge zu verstehen.

Positive Effekte auf die zu erwartenden Uplifts zeigten die Testergebnisse, wenn diese von Tests stammten, welche speziell für mobile Endgeräte entwickelt wurden. Tests dieser Kategorie zeigten in der Studie zusätzlich eine Tendenz dazu, generell häufiger erfolgreich zu sein. Berücksichtigt man bei der Bewertung dieser Ergebnisse weitere Studien, laut denen mobiler Traffic zumeist eine niedrigere Conversion-Rate als Desktop-Traffic erzielt (vgl. IntelliAd Media, 2017, S. 4) und der generelle Traffic-Anteil von mobilen Geräten im Web stetig steigt (vgl. Cisco Systems, 2017, S. 7), können die Ergebnisse dieser Arbeit als Hinweis darauf verstanden werden, dass eine Fokussierung auf die Optimierung für Nutzer mobiler Endgeräte besonders erfolgsversprechend sind.

Um die Wirkung bestimmter Hypothesen und insbesondere die Zusammenhänge der Hypothesen-Faktoren Stimulus und Intention zu untersuchen, wurden statistische Cluster gebildet, welche bestimmte Muster abbilden sollen, die besonders erfolgsversprechend oder typisch für erfolglose Split-Tests wären.

Auch wenn auf Basis der verfügbaren Daten keine statistisch belastbaren Modelle ermittelt werden konnten, zeigte die Arbeit, dass eine Clusterbildung grundsätzlich möglich erscheint und wie daraus Empfehlungen für die CRO abgeleitet werden könnten.

Mögliche Richtungen für weitere Untersuchungen könnten in der Auswahl der angewandten Faktoren für das Cluster-Modell und in der Anwendung komplexerer Algorithmen liegen. Auch eine Klassifizierung mithilfe eines Entscheidungsbaums wäre möglicherweise ein nächster Ansatzpunkt, um komplexere Zusammenhänge zu erforschen.

Anhand der ausgewählten Meta-Faktoren, wie dem Geschäftsmodell der testenden Website, der Anzahl verwendeter Stimuli im Test und dem Seitentyp, versuchte die Studie, ein logistisches Regressionsmodell zu bilden, anhand dessen sich die Erfolgswahrscheinlichkeit eines Tests unter Berücksichtigung aller Faktoren vorhersagen lässt.

Unter den Einschränkungen der Stichprobenzusammenstellung, wie in Abschnitt 4.5 beschrieben, konnte die Arbeit zeigen, dass die gewählten Faktoren prinzipiell zu einem Vorhersagemodell genutzt werden können. Anhand einiger Trainingsdaten und eines kleinen Satzes Testdaten wurde ein Testlauf durchgeführt, der das Potenzial dieser Herangehensweise bestätigt. Gerade in Anbetracht der praktischen Anwendungsmöglichkeiten werden weitere Forschungen in dieser Richtung mit den Ergebnissen dieser Studie empfohlen.

Abschließend kann zusammengefasst werden, dass alle Ziele dieser Arbeit ganz oder zumindest in Teilen erreicht wurden. Sie liefert praktische Hinweise für Unternehmen, welche Faktoren dazu beitragen können, die eigenen CRO Prozesse effektiv zu gestalten, und welche Faktoren die besten Testergebnisse begünstigen.



# LITERATURVERZEICHNIS

- Albers, Sönke (Hrsg.) (2007): Methodik der empirischen Forschung, Wiesbaden: Springer Fachmedien, 2007
- Ash, Tim/Ginty, Maura/Page, Rich (2012): Landing Page Optimization: The Definitive Guide to Testing and Tuning for Conversions, 2. Aufl., Chichester: Wiley, 2012
- Balzert, Helmut u. a. (2010): Wissenschaftliches Arbeiten: Wissenschaft, Quellen, Artefakte, Organisation, Präsentation, Herdecke: W3L-Verl., 2010
- Beck, Hanno (2014): Behavioral Economics: Eine Einführung, Wiesbaden: Springer Gabler, 2014
- Beilharz, Felix u. a. (Hrsg.) (2017): Der Online Marketing Manager: Handbuch für die Praxis, Heidelberg: O'Reilly, 2017
- Beschnitt, Martin (2009): Optimierung als Prozess: Usability Engineering, in: Frank Reese (Hrsg.), Website-Testing: Conversion Optimierung für Landing Pages und Online-Angebote - Standardausgabe, 2009, S. 27–34
- Bitkom e.V. (Hrsg.) (Juli 2017): Marketingbudgets und -maßnahmen in der ITK-Branche: Ergebnispräsentation der Unternehmensbefragung 2017, <<https://www.bitkom.org/noindex/Publikationen/2017/Studien/2017/Marketingbudgets-2017/20170728-Marketingbudget-2017.pdf>> [Zugriff 2018-08-01]
- Brandstätter, Veronika u. a. (2018): Motivation und Emotion: Allgemeine Psychologie für Bachelor, Berlin, Heidelberg: Springer Berlin Heidelberg, 2018
- Buxmann, Peter (o.J.): IKT-Branche, <<http://www.enzyklopaedie-der-wirtschaftsinformatik.de/lexikon/uebergreifendes/Kontext-und-Grundlagen/Markt/IT-Branche/index.html>> [Zugriff 2018-08-27]
- Chatterjee, Samir (Hrsg.) (2009): Proceedings of the 4th International Conference on Persuasive Technology, New York, NY: ACM, 2009
- Cisco Systems (Hrsg.) (2017): Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2016–2021, <<https://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/mobile-white-paper-c11-520862.pdf>> [Zugriff 2018-08-25]
- Deutsche Post AG (Hrsg.) (Juni 2018): Werbemarkt Deutschland: Dialogmarketing.Monitor 2018 / Studie 30, <[https://www.deutschepost.de/content/dam/dpag/images/D\\_d/dialog-marketing-monitor/dp-dmm-monitor-2018.pdf](https://www.deutschepost.de/content/dam/dpag/images/D_d/dialog-marketing-monitor/dp-dmm-monitor-2018.pdf)> [Zugriff 2018-07-20]

Düweke, Esther/Rabsch, Stefan (2011): Erfolgreiche Websites: SEO, SEM, Online-Marketing, Usability, 2011

Econsultancy/Adobe Systems (Hrsg.) (März 2018): Digital Intelligence Briefing: Digitale Trends 2018, <[https://www.adobe.com/de/modal-offers/econsultancy\\_digital\\_trends\\_2018\\_report.html](https://www.adobe.com/de/modal-offers/econsultancy_digital_trends_2018_report.html)>

Fahrmeir, Ludwig u. a. (2003): Statistik: Der Weg zur Datenanalyse, Berlin, Heidelberg/s.l.: Springer Berlin Heidelberg, 2003

Fogg, B. J. (2009): A behavior model for persuasive design, in: Samir Chatterjee (Hrsg.), Proceedings of the 4th International Conference on Persuasive Technology, 2009, S. 1, <https://doi.org/10.1145/1541948.1541999#>

Forrester Research Inc. (Hrsg.) (2018): Forrester Opportunity Snapshot: AB Tasty: Maximize customer value through strategic Conversion Rate Optimization, <<https://www.abtasty.com/ebooks/cro-forrester-research/>>

Foscht, Thomas/Swoboda, Bernhard/Schramm-Klein, Hanna (2015): Käuferverhalten: Grundlagen - Perspektiven - Anwendungen, 5. Aufl., Wiesbaden: Springer Gabler, 2015

Goward, Chris (2013): You should test that!: Conversion optimization for more leads, sales and profit or the art and science of optimized marketing, Hoboken, NJ: Wiley Sybex, 2013  
– (09.03.2018): The LIFT Model: Use these six factors to increase your conversion rate, <The LIFT Model: Use these six factors to increase your conversion rate> [Zugriff 2018-08-24]

Heinemann, Gerrit/Haug, Kathrin/Gehrckens, Mathias (Hrsg.) (2013): Digitalisierung des Handels mit ePace, Wiesbaden: Springer Fachmedien Wiesbaden, 2013

intelliAd Media (Hrsg.) (2017): intelliAd E-Commerce Branchenindex Q3/2017, <[https://www.intelliad.de/wp-content/uploads/2017/11/2017\\_intelliAd\\_Studie\\_B Branchenindex-Q3-2017.pdf](https://www.intelliad.de/wp-content/uploads/2017/11/2017_intelliAd_Studie_B Branchenindex-Q3-2017.pdf)> [Zugriff 2018-07-04]

Jacobsen, Jens (2011): Website-Konzeption: Erfolgreiche Websites planen, umsetzen und betreiben, 6. Aufl., München: Addison-Wesley, 2011

Kähler, Wolf-Michael (2002): Statistische Datenanalyse: Verfahren verstehen und mit SPSS gekonnt einsetzen, 2. Aufl., Braunschweig: Vieweg, 2002

Kattau, Nils (2017): Conversion-Optimierung, in: Felix Beilharz/Nils Kattau/Karl Kratz/Olaf Kopp/Anke Probst/Guido Pelzer/Markus Kellermann/Manuela Meier/Ingo Kamps/Markus Vollmert/Niklas Plutte/Wolfgang Neider (Hrsg.), Der Online Marketing Manager: Handbuch für die Praxis, 2017, S. 65–101

- Kaushik, Avinash (2010): Web analytics 2.0: The art of online accountability & science of customer centricity, Hoboken, NJ: Wiley, 2010
- Kollmann, Tobias (2011): E-Business: Grundlagen elektronischer Geschäftsprozesse in der Net Economy, 4. Aufl., Wiesbaden: Gabler, 2011
- Kreutzer, Ralf T. (2014): Praxisorientiertes Online-Marketing: Konzepte - Instrumente - Checklisten, 2. Aufl., 2014
- Kroeber-Riel, Werner/Gröppel-Klein, Andrea (2013): Konsumentenverhalten, 10. Aufl., München: Vahlen, 2013
- Krüger, Jörg Dennis (2011): Conversion Boosting mit Website Testing, Heidelberg u. a.: Verlagsgruppe Hüthig Jehle Rehm, 2011
- Lauber, Dirk (2013): E-Commerce an der Schwelle zur Sättigungsphase: Produktivität von E-Commerce-Aktivitäten wird erfolgskritisch, in: Gerrit Heinemann/Kathrin Haug/Mathias Gehrckens (Hrsg.), Digitalisierung des Handels mit ePace, 2013, S. 105–122
- Martínez-López, Francisco J. (Hrsg.) (2014): Handbook of Strategic e-Business Management, Berlin, Heidelberg/s.l.: Springer Berlin Heidelberg, 2014
- Melzer, Almut (2015): Six Sigma - Kompakt und praxisnah: Prozessverbesserung effizient und erfolgreich implementieren, 2015
- Morys, André (2011): Conversion-Optimierung: Praxismethoden für mehr Markterfolg im Web, Frankfurt am Main: entwickler.press, 2011
- Reese, Frank (2009): Die Kunst des Experiments: Oder:Wozu eigentlich testen, in: Frank Reese (Hrsg.), Website-Testing: Conversion Optimierung für Landing Pages und Online-Angebote - Standardausgabe, 2009, S. 13–26– (Hrsg.) (2009): Website-Testing: Conversion Optimierung für Landing Pages und Online-Angebote - Standardausgabe, Göttingen: BusinessVillage, 2009
- Rohrlack, Christian (2007): Logistische und Ordinale Regression, in: Sönke Albers (Hrsg.), Methodik der empirischen Forschung, 2007, S. 237–282
- Sharma, Himanshu (2015): Maths and stats for web analytics and conversion optimization: The journey from ‚data driven‘ to ‚data smart‘, USA, 2015
- Wiedenbeck, Michael/Züll, Cornelia (2010): Clusteranalyse, in: Christof Wolf/Henning Best (Hrsg.), Handbuch der sozialwissenschaftlichen Datenanalyse, 2010, S. 524–553
- Wirtz, Bernd W. (2016): Electronic Business, 5. Aufl., 2016

Wolf, Christof/Best, Henning (Hrsg.) (2010): Handbuch der sozialwissenschaftlichen Datenanalyse, 2010

Wolf, Talia (2018): 90% of our decision making is irrational, <<https://www.conversioner.com/blog/90-of-our-decision-making-is-irrational>> [Zugriff 2018-08-13]

Zerdick, Axel (2001): Die Internet-Ökonomie: Strategien für die digitale Wirtschaft, 3. Aufl., Berlin: Springer, 2001