

## The reliable path to success

Using A/B & Multivariant
Testing for Onsite Optimization

75

"Every online shop has its own unique experience with customers and their buying behavior. Nevertheless, hardly anyone can permanently predict which variants of the shop or the marketing approach are the best. The Users' behavior and their needs are difficult to generalize – there is no such thing as a "John Doe" online. To ensure a positive user experience for everyone, actions should be validated through testing and iteratively improved."





### ABOUT TRBO \_\_\_\_

trbo is the leading technology provider for dynamic onsite personalization, optimization and testing. trbo's AI-driven onsite personalization platform allows users to customize website content, A/B and multivariant test, and serve recommendations down to the individual level using machine learning. trbo's self-learning algorithm analyzes user behavior based on 50+ visitor characteristics, creating highly personalized, unforgettable website experiences for their customers. Incredibly agile and easy to use, trbo's toolset allows users to make changes in real time and serve their customers' needs right away.

Major global brands across all industries such as L'Oreál, WeightWatchers, Eddie Bauer, Ecco, Zwilling, Porsche Design and many more rely on trbo's comprehensive services.

trbo was founded in 2013 by CEO Felix Schirl with global headquarters in Munich, Germany and US headquarters in Miami, Florida.

### INTRODUCTION \_

A/B and multivariant tests are the most important tool for optimizing key performance indicators (KPIs) such as the click-through rate or the conversion rate in the best possible way – by improving the usability of and user experience on the website. They allow checking the success of onsite functionalities and campaigns and to draw reliable conclusions.

#### **TESTING - IMPORTANT TOOL WITH IMMENSE BENEFITS**

Businesses always want to offer their users the best shopping and surfing experience. But which actions are real added value – and which ones can users do without? A/B and multivariant tests provide answers to these questions.

Nevertheless, many companies still don't use tests at all or not consistently. This is actually surprising, given that they are one of the most important means of conversion rate optimization. And there is still a lot of room for improvement here. Depending on the industry, the average conversion rate in Germany was 0.4% (travel) to 10% (online pharmacies). Globally, the average conversion rate in eCommerce was between 1.5% and just under three percent.¹ In contrast, there is the 92:1 rule: 92 dollars are spent to acquire new customers. But only \$1 of that goes to conversion optimization.² The takeaway? The strategy of using high budgets to boost traffic on

the site will fall short in the long term.Instead, users should be offered a special experience on the website, shopping should be made as easy and convenient as possible for them and they should always be offered the best service. The more optimized and personalized the website is, the more comfortable users feel and are willing to shop. Accordingly, the conversion rate as well as sales will rise. But how can this be achieved? And which actions really resonate with users? A/B and multivariant tests provide the answers to these questions: quickly and scientifically valid. Thus, optimization is no longer based on gut feeling, but on reliable figures and results.

As a leading provider for onsite personalization, optimization and testing, trbo helps numerous online retailers, such as Telefónica, Zwilling and L'Oréal, to continuously improve and test the user experience on their website.

<sup>1</sup> https://www.irpcommerce.com/en/gb/ecommercemarketdata.aspx https://de.statista.com/statistik/daten/studie/677869/umfrage/conversion-rate-nach-branchen/

<sup>2</sup> https://research.aimultiple.com/cro-stats/

# 01

# THE DIFFERENCE BETWEEN A/B AND MULTIVARIANT TESTING

Testing methods are used to iteratively improve the performance of a company's own website or app regarding certain KPIs. There are different types of tests that can be used here.

The most common one is the A/B test, also called split test. It involves testing two versions (A and B) against each other. Version A contains the control group (usually the original version of the website) and version B the new variation to be tested. The possibilities for changes are manifold: changing single elements and links or whole parts of pages can be tested. The users are randomly distributed to the two versions. The shop is free to choose how many users should see each of the two versions. The most common distribution is 50:50.

Multivariant testing, also known as A/B/n testing, is the extended form of the frequently used A/B test. Classically, a multivariant test primarily refers to the adaptation of a variant into several presentation options. In multivariant or A/B/n tests, the variable "n" stands for the unlimited number of possible variants that can be tested against each other at the same time. For example, a newsletter pop-up: in this case, it could be tested whether the form works best with first and last name as a required field, a voluntary field, or completely without providing the name. The users are randomly assigned to one of the

different variants, just like in a regular A/B test. However, this type of test requires a larger number of users in order to achieve a meaningful result.

The disadvantage of the multivariant or A/B/n test: the analysis is based on an average customer. However, this cannot be defined in a generally valid way. Also, other deviations of the user cannot be taken into account. For example, regular customers or big spenders can falsify the results of the test due to their buying behavior, which differs greatly from the norm. Ideally, therefore, different testing groups are formed, which can also be used to evaluate devices separately from one another. An extension of addressing different target groups is the implementation of dynamic A/B/n tests. In this case, the variants are adapted repeatedly on the basis of user behavior.

A special category in testing is the so-called multi-armed bandit testing. This is a complex form of testing that uses machine learning and AI to dynamically optimize the tests. Then, the traffic is dynamically directed to the best performing variants – allowing the variants with the best performance to be identified more quickly and determined to be statistically significant. Otherwise, if there is not enough traffic, reaching statistical significance on all variants may take a long time.

# 02

### TESTING – HOW DO YOU ENSURE A SUCCESSFUL START?

How do you successfully start testing and what should you think about beforehand? What figures and terms do you need to know and is there a recommended procedure? We will clarify these questions in the next chapter.

Smaller changes to the website, such as the replacement of a headline text or a bug fix, usually do not require extensive testing. However, even for changes to call-to-actions (CTA) – for example, the color of the button or the CTA text – it pays off to set up a test. Page elements that users interact with basically have big potential for improvement – but they also lead to poor decisions made on gut instinct. The same applies to the traffic on the page. If the number of visitors is too low, there may not be any significant results at the end of the test.

**DEVELOPING A HYPOTHESIS** 

Designing a proper test is time-consuming. But simply setting up a test without giving it a second thought and not being able to derive any meaningful results takes up even more time. Accordingly, shops should properly organize their test setup. The two important concepts in A/B testing are the null hypothesis and the alternative hypothesis. The null hypothesis attempts to disprove or confirm the assumption that there is no correlation between the original and the alternative variant, i.e. that the alternative does

not result in a better conversion rate. The concept of the alternative hypothesis is based on the assumption that there is a dependency and one wants to confirm or refute this very hypothesis. In an A/B test, both hypotheses are tested against each other. They are therefore connected to each other like yin and yang. Usually, the goal of an A/B test is to reject the null hypothesis and accept the alternative hypothesis. Basically, you want to prove that the alternative is more likely to deliver more conversions. Important questions should also be included when formulating the hypothesis that serves as the basis for a successful A/B test: Which target group do you want to reach, which parts of the shop do you want to adjust and, most importantly, which KPI do you want to optimize? What exactly is to be achieved, what do you suspect is going wrong so far?



An example of such a hypothesis for an A/B test: The bounce rate for SEA entries is high. The hypothesis could be: "Our shop visitors lack possible alternatives for SEA entries on product detail pages, which is why they quickly leave the shop again if they do not like the product. By enriching the page with more alternative products, the bounce rate will drop by X percent."

UNDERSTANDING SIGNIFICANCE & CONFIDENCE

Anyone working with tests should be familiar with the terms "significance" and "confidence". After all, they are important for accurately evaluating a test.

Confidence refers to the percentage of probability that a certain statement will be true. In practice, the recommended confidence level is usually 95 percent. For example, if variant B (using alternative products) improves conversion and lowers the bounce rate, a confidence level of 95 percent would mean that B would work better with a certainty of 95 percent.

Significance, in turn, indicates the point at which the measured correlation between the variables does not occur by chance and thus the result can be applied to the entirety. For this purpose, a probability of error is defined (p-value). Its upper limit is indicated by the significance level. Assuming a confidence level of 95 percent, the significance level is 5 percent (100% minus confidence level of 95%).

If one defines a confidence level of 95%, the probability of error (p-value) is max. 5% (100% minus 95%). If the test result is thus below 5%, the

hypothesis can be accepted and is referred to as significant.

The significance level is the probability of rejecting the null hypothesis when it is true. In other words, a significance level of 0.05 indicates a 5% risk (100% minus confidence level of 95%) of concluding that a difference exists when there is no actual difference.

Sounds complicated? Well, it is. Fortunately, many testing tools, such as trbo, include a significance calculator. This way, shops can always check whether the campaign can be evaluated.



If a test takes an extended period of time to become significant, it may also indicate that the element being tested has no real impact. In this case, the test should be ended and restarted in a different form.

### PLANNING & STRATEGY: KPIS, DURATION, GOAL SETTING

Which KPIs do we want to check and when should we start testing? These are two questions you need to ask yourself before starting the testing process.

#### Select duration and timing of the test

Finding the right time is important, as seasonal factors (e.g. Christmas) or the purchasing power of users can lead to fluctuating purchase and traffic figures. Accordingly, you should schedule tests in "purchasing-neutral" months that do not include holidays such as Easter and Christmas (unless you want to test elements that are only used at this time). If, for example, the aim is to increase purchases, these periods may deliver false positive results.

The duration of the test is also essential. This is very shop-specific, as how quickly a test becomes significant depends strongly on the traffic. The expected uplift/downlift and the number of users can be used to determine in advance how long a test will take – and to check whether a test setup is really worthwhile. Test duration calculators are available for this purpose. The smaller the size of the groups, the longer the test will take and the less likely a significant result will be.

It is important to never end the test too soon, as results often change with increasing user numbers. If you end testing too early (either because of positive or negative results), you run the risk of drawing wrong conclusions, as the numbers might have again changed completely in the course of the test.



#### Do not test too much at once

Using A/B testing, you can quickly and easily find outwhether a specific setup works or not. However, not every change to an element should be tested without restriction – at least not simultaneously. Especially when considering sales and conversion rates, you should if possible conduct only one test per (sub-)page. Otherwise, the tests may influence each other and correlations that do not exist may be established. If the pages are clearly separated from each other – such as the homepage and category page – several tests may run at the same time, of course. However, you should not run five tests on the homepage.

Furthermore, it is important to not only consider the page itself, but also the flow of users. If, for example, numerous links on the homepage lead to the category page, and both pages are running a test in which the conversions serve as the primary KPI, they can interfere with each other. If a test is conducted in the checkout area, no other page should be tested for conversions, as the user will inevitably encounter the test in the checkout area.

#### Selecting the right users

Not every test is suitable for all users, some users can even significantly falsify the test result. A quick example: If you want to test the effects of a new wording on newsletter registrations, customers who have already registered should be excluded from this test. The segmentation of users is therefore of great importance.

#### Choosing the right KPIs

As there are countless possibilities for testing, the first step should always be determining the relevant KPIs for evaluating the test result. A distinction should also be made between a primary (decisive) KPI and secondary KPIs. Frequently used measurement values that

can be used as primary KPIs are, for example, the click through rate (CTR), the conversion rate or the conversion value and user value. An example of a secondary KPI is measuring the achievement of a defined funnel step to find out even more about the impact of the test on user behavior.



# TESTING WITH PURPOSE & UNDERSTANDING

#### **DEFINE HYPOTHESIS**

For successful testing, clear objectives and hypotheses must be defined.

#### PLAN STRATEGICALLY

Tests with the highest potential impact should be given high priority – run the most important tests first.

#### STATISTICALLY VALID TESTING

Consider significance and confidence and do not end your tests too early.

Only those who plan and implement tests properly will obtain reliable results.

What are the benefits?

#03

## A/B AND AND MULTIVARIANT TESTS IN ACTION

But what do meaningful tests look like? At trbo, we develop exciting test setups with our customers on a daily basis to validate the benefits

of personalization and optimization. We have compiled a small selection of ideas in the following chapter.

#### **DESIGN AND LAYOUT OF PRODUCT DETAIL PAGES**

Clear and detailed information

In eCommerce, you need to pay special attention to the design of product detail pages – after all, this is where the decision to purchase or not mostly takes place. Therefore, these pages should be as informative as necessary, but also have a clear layout.

How detailed product texts should be or whether an expandable teaser is sufficient can be quickly determined using A/B tests. Which services provide added value (free shipping, free returns, etc.) and where should these best be placed? All of this can be tested.

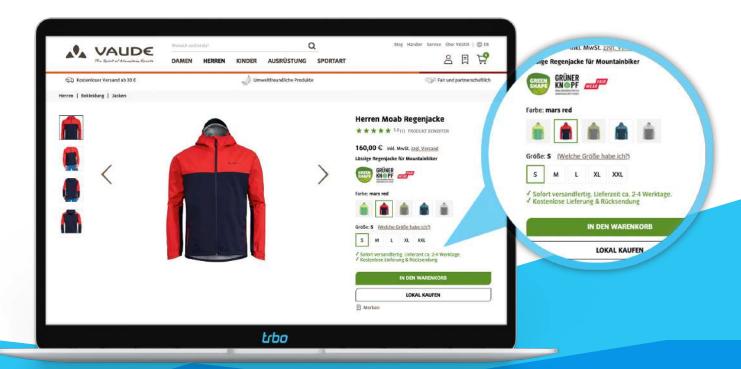
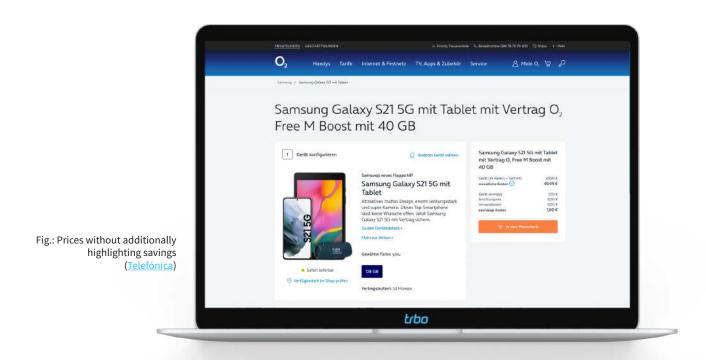


Fig.: Fade-in of information about the product's availability as well as shipping and return conditions (VAUDE)

#### Which arguments for purchase are there?

Are users convinced by the price or does the opinion of others influence more? Many users are more likely to buy if they see honest product reviews. Other users are more price-sensitive and can be convinced primarily by strike prices or the display of their savings (in percent). Where should product reviews be placed and in what form (via asterisks or as direct quotes from real customers)? Does displaying a strike price in combination with the amount saved make more sense? Find out with A/B and multivariant tests.



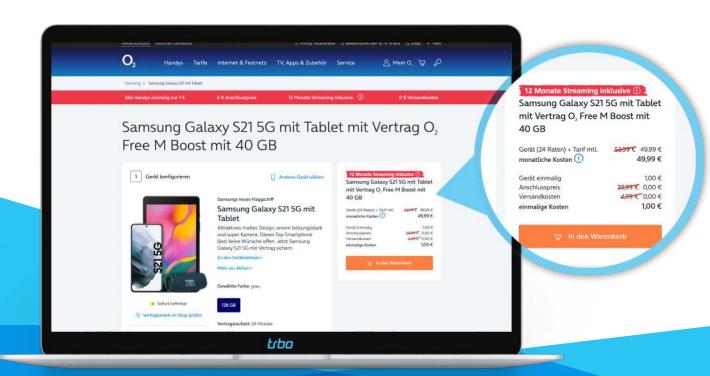


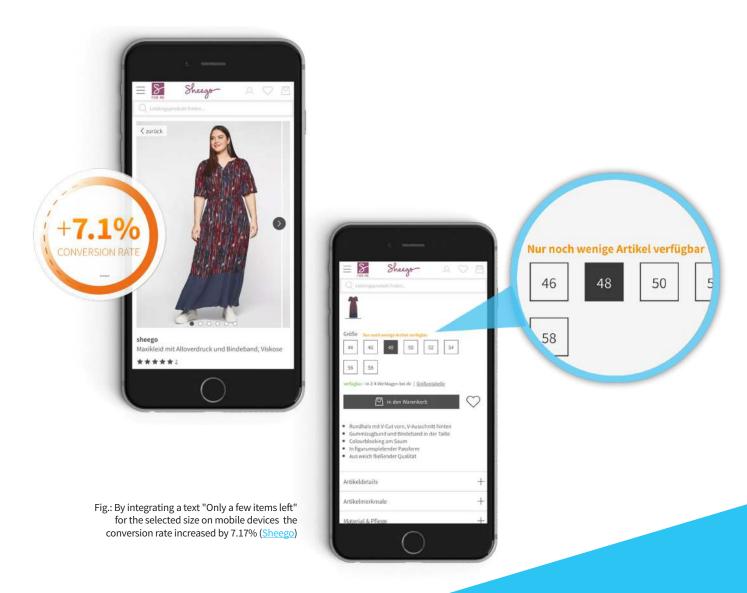
Fig.: Additional integration of strike prices(Telefónica)

#### Exerting pressure?

Do countdowns of discounts lead to a faster conclusion of a purchase? Does the community have an impact and drive users to a purchase more quickly? Does it make sense to highlight the (limited) availability of products once again?

To increase the pressure to buy, there are various possibilities. These include elements such as progress bars that keep filling up towards free shipping. Countdowns are also a popular way to time discounts and guide users to complete a purchase more quickly. Group Motivation elements show how many users have recently viewed and/or purchased a product. Availability displays also lead to a quicker purchase decision, as the limited stock catches the user's eye.

Do these elements also work in your shop? You can find out quickly and easily by testing.

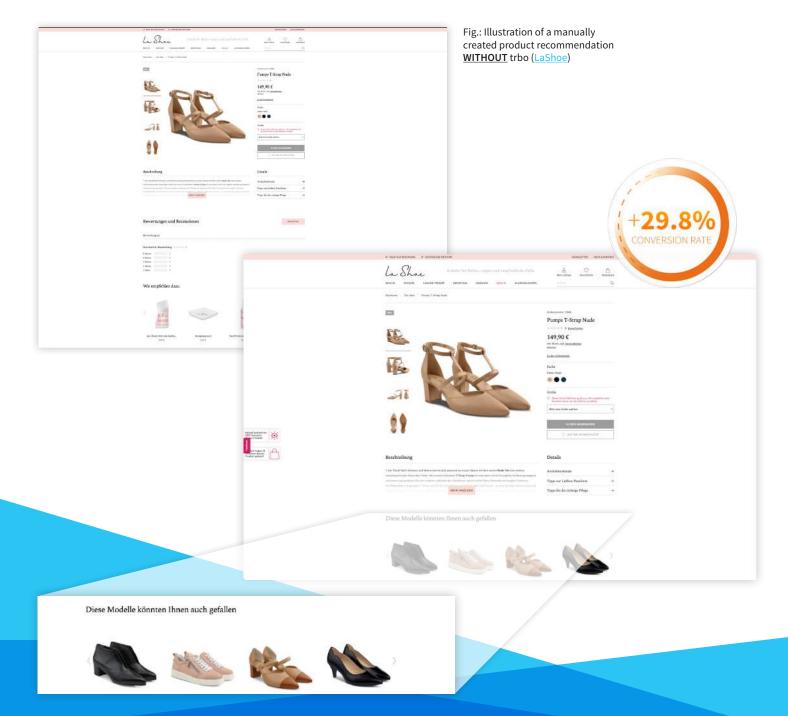


#### MEANINGFUL PRODUCT RECOMMENDATIONS

#### What kind of recommendations?

There are all kinds of product recommendations and underlying display logics. These include, for example, top sellers of the shop, recommendations of similar products, manually assigned recommendations (e.g. accessories), "Other customers also bought..."-recommendations and personalized recommendations based on surfing and buying behavior – to name just a few possibilities.

Which logic promises the best results and where the product recommendations should be placed (at the top of the homepage or further down, on product detail pages directly below the product or below the ratings, etc.) can be statistically validly evaluated with A/B and multivariant tests.



 $Fig.: The integration of the trbo-recommendation leads to a conversion rate uplift of 29,8\% \ (\underline{LaShoe})$ 

#### Would you like to buy some more?

How can you exploit additional potential from product recommendations? Shops can make use of various cross-sell and upsell mechanisms. Recommendations in the shopping cart lead to additional purchases.

The aim of upselling is to ensure that the customer spends more money than he had originally planned. In order to achieve that, shops try to convince the customers that a higher-quality and therefore often more expensive product is better suited to their needs than the item they have been looking at so far.

Using cross-selling means suggesting additional products to customers that could match the item they have already selected. For example, matching belts are also offered along with the jeans. This works in the shopping cart as well – often similar to the checkout areas of brick-and-mortar stores, where customers often buy additional products right before the checkout. In the online shopping cart, additional products can also be displayed.

With testing, shops can evaluate which product logics and placements offer additional sales potential through recommendations.

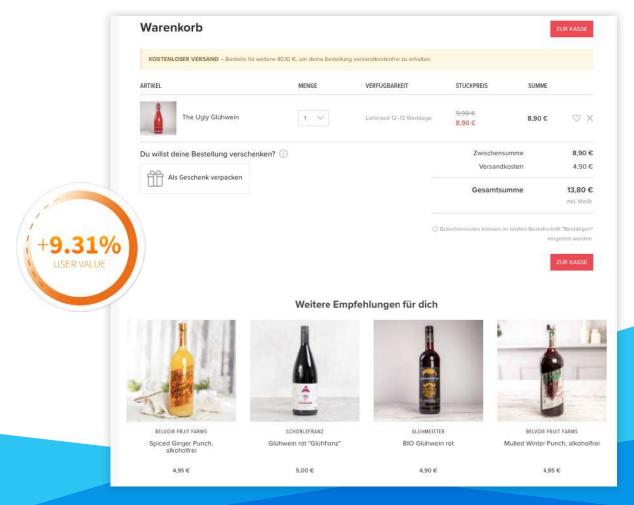


Fig.: Additional product recommendations in the shopping cart increased the conversion rate by 8.2%, the user value increased by 9,31% (Foodist)

#### **OPTIMIZING THE SHOP'S NAVIGATION**

Another important element for optimization is improving the navigation. After all, users will only be directed to the right product and buy if they can find their way around the page.

#### How do users search?

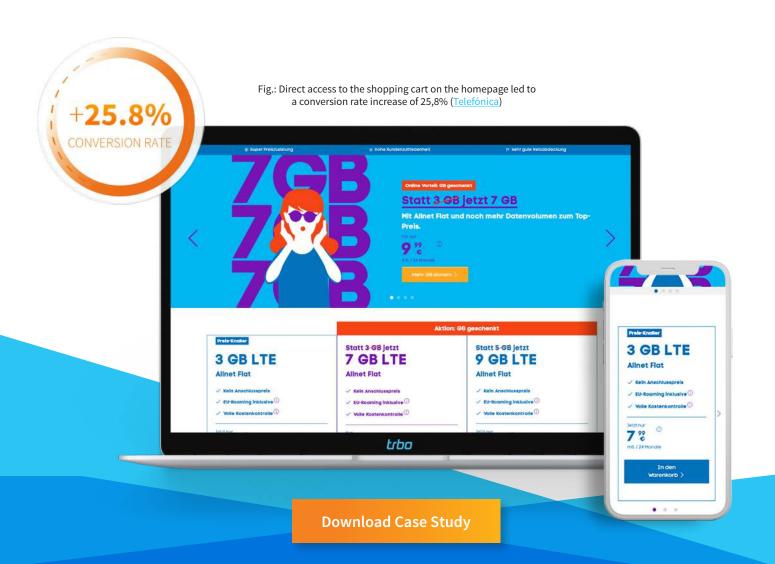
During optimization, the search bar should not be forgotten as it ideally quickly leads to the target. The layout alone can cause changes in search behavior. Do users respond better to a prominent search slot? Does it make sense to directly display product recommendations in the search or do users prefer to see results as lists?

#### New or changed menu items?

Bargain hunters are often looking for the best offers – therefore the sale menu item should be displayed much more prominently. Does the shopping experience improve if the menu items are sorted individually according to user behavior? You should definitely test this!

#### Directly to the shopping cart?

Sometimes the easiest way leads directly to the goal. Perhaps users don't want to find out about the various products in detail, but would rather be taken directly to the shopping cart.



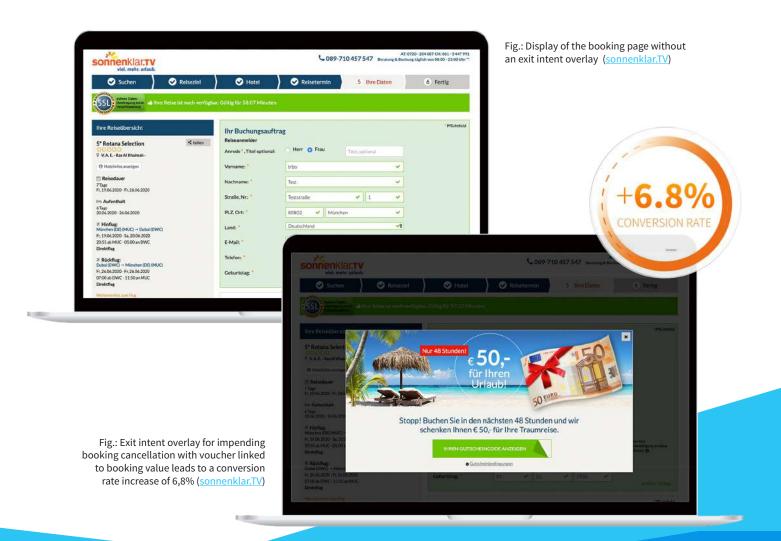
#### AVOID SHOPPING CART ABANDONMENT & BRING USERS BACK INTO THE BUYING PROCESS

Do you really want to leave already?

The term exit intent describes the point at which a user threatens to bounce off a website. Modern exit intent technologies help prevent users from leaving the website by displaying overlays. If the mouse pointer moves towards the browser's address bar or the close icon in the upper left corner, an overlay appears. This overlay can be placed in the middle or on the side to display a message to the user and thus trigger a specific action.

In this case, the goal is to keep the potential customer on the page and, in the best case, persuade him to make a purchase. The content of such overlays can be diverse: products in the shopping cart, further product recommendations, coupons or a newsletter registration. With the exit intent method, the conversion and click-through rate of the website can be further optimized and the bounce rate or shopping cart abandonment can be reduced.

Which type of exit intent overlay is suitable and how do users react to it? You can test this using A/B and multivariant tests.



**Download Case Study** 

#### **HOMEPAGE PERSONALIZATION**

How does a deeper personalization of the home page affect users?

The homepage can be referred to as the store window of the online shop. As in brick-and-mortar retail, it gives potential customers a first glimpse of the product range and points out current promotions. Based on the homepage, users decide within a very short time whether the webshop meets their expectations. The dynamic, personalized homepage is therefore an indispensable tool for online shops. There are numerous ways in which a personalized homepage can be created to meet the expectations of users.

One possibility is to dynamically exchange teaser areas based on the user's interests, as well as to display (personalized) product recommendations. Returning users can be brought back into the buying process by integrating their most recently viewed products.

What type of personalization do users like best? Are there user groups that respond particularly well to personalization? Which combination is most promising and where should the product recommendations be seen?

All this can be tested with A/B and multivariant tests.

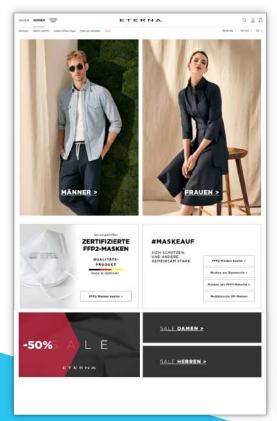
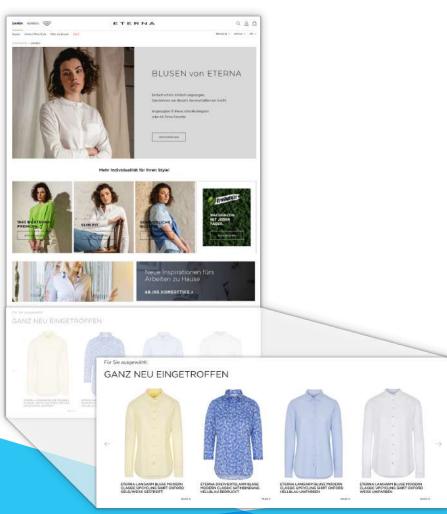


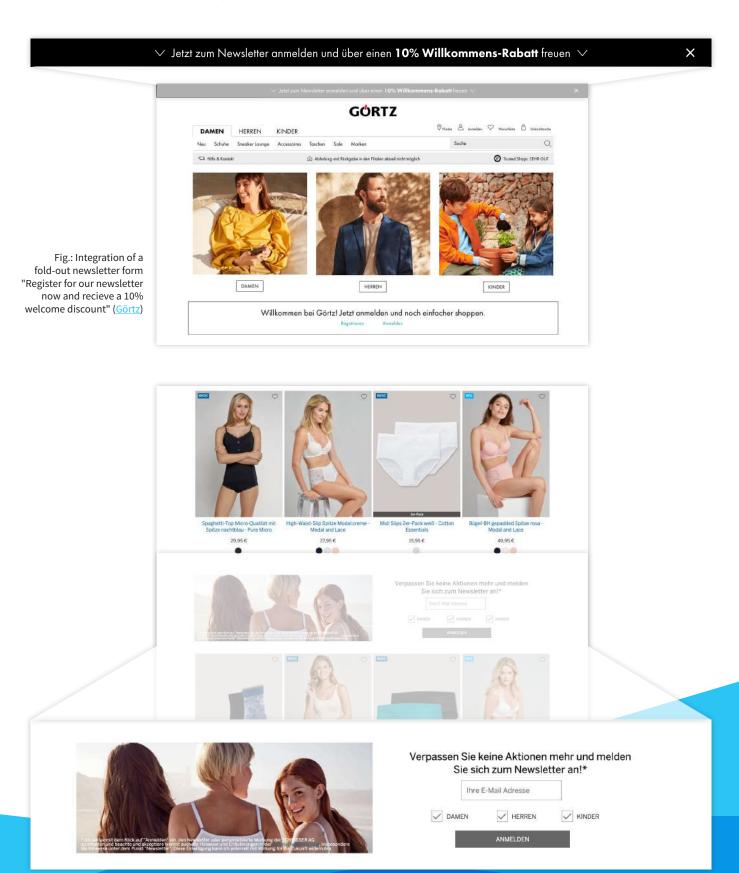
Fig.: Left: Default homepage, Right: Personalized homepage according to interest in women's fashion with customization of teaser areas and personalized recommendations (ETERNA)



#### **GENERATE LEADS**

Is the fear of the pop up justified?

Newsletters are an important means of customer retention. But how do you gain more subscribers without alienating users? Are exit intent overlays really annoying or are they very well received by users? This can be found out in an A/B or multivariant test.



# 04

## A/B AND MULTIVARIANT TESTING – EVALUATION & LEARNINGS

You have now successfully identified various elements in your shop, evaluated them in terms of feasibility and scheduled the test. After a few days or weeks, the test is significant and it is time to evaluate the results.

te the results.

Funnel Reporting

Evaluation of the results

Professionelle Plattformen zur Professional platforms for personalization and optimization offer the possibility to evaluate tests directly within the tool. In the dashboard, uplifts (or downlifts) can then be read quickly and clearly – sorted according to the most important KPIs. An evaluation in the leading tracking tool enables a detailed analysis that, for example, also considers other factors that cannot be integrated in the test setup alone. Therefore, test events should always be transmitted to the leading analysis tool as well.

When you launch campaigns and actions on your website – for example with trbo – you usually want to know how and when they influence user behavior. For such an evaluation, it comes in handy to use funnel reporting to view the relevant funnel on the website. In this way, you can see at which point in the customer journey the campaign influences behavior. Depending on which primary KPI is used for the test - for example, accessing the shopping cart - it is possible to check whether and how the campaign also influences further steps and whether the effect can be significantly attributed to this campaign.

This way, the tool extrapolates the results and

possible distortions due to increased traffic on

one variant are compensated.



When evaluating tests in which the traffic was distributed very unevenly (e.g. 70/30), you should ensure an equal distribution in the evaluation.

Generally, a funnel reporting can be applied to any campaign, whether A/B or multivariant testing, changes without testing, or more complex measures. Based on statistics, it is possible to determine how many people interacted with the action and by what percentage a downlift or uplift of the respective campaign goal was achieved. For example, if the goal is to increase the number of users in the shopping cart, it is possible to identify how the campaign affected this.

The following example examines the effect of a shopping cart direct entry via the category page. Instead of being redirected to the product detail page, the test group was shown the product details directly on the category page. All links led the users directly to the shopping cart. The example shows: The direct link to the shopping cart led to a significant uplift in all further funnel steps.

### PERSONALIZATION AND A/B/N TESTING SHOULD GO HAND IN HAND

Personalization and A/B or A/B/n testing should always go hand in hand. Future trends are moving towards dynamic optimization of tests: there will no longer be just one winner of a test, there will be many winners for different target groups. While the overall winner can still be identified and evaluated, the way to go in testing is evolving. Groups should be illuminated based on performance.

#### Every test brings a result

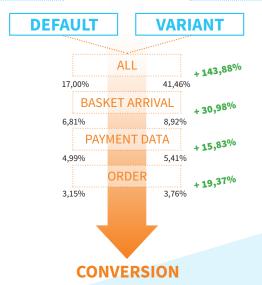
Not always do the tested setups lead to huge uplifts – sometimes they even aim at the opposite. But even that is relevant, if only to determine that the action is not well received by users. Thus, a negative test does not directly mean failure, but represents the starting point for further optimization. Perhaps you simply addressed the wrong users with the action? Then test it on another segment!

#### Establish a testing culture

Testing and conversion optimization are no sprints, but marathons. Just because a test was successful doesn't mean that you can sit back and set up all your campaigns according to the same pattern. Successful companies always continue testing – because user behavior changes and what was once well received does not necessarily continue to work.

- Overview page
- Product details on PDP
- All links lead to the PDP

- Overview page
- Product details on overview page
- All links lead directly to the shopping cart



#### **Observations and results**

- Very strong uplift among users in the shopping cartStrong improvement in all further funnel steps
- → A lot of unqualified traffic, which however brings a significant uplift in total

### # CONCLUSION

# A/B AND MULTIVARIANT TESTS – AN ONGOING SUCCESS

Even though there are many pitfalls with A/B testing, the biggest mistake is not testing at all. Relying solely on gut feeling may work from time to time, as no one knows the shop as well as the shop owner himself. But in order to really prove your point and check whether the use of a tool is really worthwhile, you cannot ignore A/B tests. It's amazing what effect even small adjustments can have that you did not expect. And even if you think you know your shop, it is a very subjective perception whether adjustments are good or bad.

With a few simple rules and the help of the right

tool, even elaborate test setups can become a piece of cake:

- → Define clear objectives before setting up the test
- → Set up a hypothesis that needs to be proven or disproven
- → Be aware of the testing period (seasonality, advertising campaigns, etc.)
- → Set up tests one after the other in order to avoid interactions
- → Consider the significance of the test
- → Never end tests too early

#### USING TRBO GIVES YOU THE FOLLOWING ADVANTAGES

- With trbo, you can use all actions for onsite personalization, optimization and testing and always have 100% control over the success of your campaigns on our onsite personalization platform.
- The simple integration of trbo technology reduces your time should further questions arise, your
  personal account manager is available to help you implement your campaigns in the best possible
  way.
- The more data, the more comprehensive the view of each customer. trbo enables data to be harnessed, enriched, and merged across all channels, making 360-degree personalization reality.
- trbo collects over 50 visitor characteristics in real time and fully automated. This data can be used Al-based to improve the browsing and shopping experience on your website.
- Use all testing options in trbo (A/B- and multivariant tests). An integrated significance calculator
  provides a quick overview for the evaluation of the campaigns.



conversion rates and sales? Feel free to contact us!

We are happy to present how your webshop can benefit

Arrange your free and noncommital demo appointment now.

#### trbo Inc.

429 Lenox Ave Miami Beach, FL 33139 Tel: 1-844-200-7222 (toll free) E-Mail: contact@trbo.com

Web: www.trbo.com