



**AgEcon** SEARCH  
RESEARCH IN AGRICULTURAL & APPLIED ECONOMICS

*The World's Largest Open Access Agricultural & Applied Economics Digital Library*

**This document is discoverable and free to researchers across the globe due to the work of AgEcon Search.**

**Help ensure our sustainability.**

Give to AgEcon Search

AgEcon Search  
<http://ageconsearch.umn.edu>  
[aesearch@umn.edu](mailto:aesearch@umn.edu)

*Papers downloaded from **AgEcon Search** may be used for non-commercial purposes and personal study only. No other use, including posting to another Internet site, is permitted without permission from the copyright owner (not AgEcon Search), or as allowed under the provisions of Fair Use, U.S. Copyright Act, Title 17 U.S.C.*

# **HOW FRAGILE IS THE CREDIBILITY OF A QUALITY LABEL? A QUASI-NATURAL EXPERIMENT USING THE EXAMPLE OF STIFTUNG WARENTEST**

Andreas Hildenbrand

Institut für Betriebslehre der Agrar- und Ernährungswirtschaft,  
Justus-Liebig-Universität, Gießen

Rainer Kühl

Institut für Betriebslehre der Agrar- und Ernährungswirtschaft,  
Justus-Liebig-Universität, Gießen

Anne Piper

Institut für Betriebslehre der Agrar- und Ernährungswirtschaft,  
Justus-Liebig-Universität, Gießen

Kontaktautor: [Anne.Piper@agr.uni-giessen.de](mailto:Anne.Piper@agr.uni-giessen.de)



Schriftlicher Beitrag anlässlich der 55. Jahrestagung der  
Gesellschaft für Wirtschafts- und Sozialwissenschaften des Landbaues e.V.  
**„Perspektiven für die Agrar- und Ernährungswirtschaft nach der Liberalisierung“**

Gießen, 23.-25. September 2015

# HOW FRAGILE IS THE CREDIBILITY OF A QUALITY LABEL?

## A QUASI-NATURAL EXPERIMENT

### USING THE EXAMPLE OF *STIFTUNG WARENTEST*

#### Abstract

In 2013, *Stiftung Warentest* tested hazelnut chocolate for their leading magazine, called *Test*. *Stiftung Warentest* is one of the most important consumer organizations in Germany. *Ritter Sport* is a high-quality producer of chocolate in Germany. Their hazelnut chocolate did not pass the test. It was given the grade of unsatisfactory. *Stiftung Warentest* accused *Ritter Sport* of labeling an artificial flavoring as a natural flavoring. *Ritter Sport* rejected the accusation. They went to court and won the trial. Using the *Ritter Sport* versus *Stiftung Warentest* case, we analyze whether negative headlines really undermine the credibility of a quality label by examining *Stiftung Warentest* and their *Test* label. In addition, we examine what can be done to restore or, more generally, increase the credibility of a quality label. Based on a quasi-natural experiment, we find that the negative headlines on *Stiftung Warentest* have undermined the credibility of the *Test* label. We also find that the credibility of the *Test* label can be restored by providing reference values to the tests, strengthening the independence of *Stiftung Warentest*, and using laboratory methods in the tests.

#### Keywords

information (D8); product quality (L1); quality label; source credibility (M3).

#### 1 Introduction

In 2013, *Stiftung Warentest* tested hazelnut chocolate for the December issue of their leading magazine, called *Test*. *Stiftung Warentest* is one of the most important consumer organizations in Germany. *Ritter Sport* is a high-quality producer of chocolate in Germany. Their hazelnut chocolate did not pass the test. It was given the grade of unsatisfactory. *Stiftung Warentest* accused *Ritter Sport* of labeling an artificial flavoring as a natural flavoring. *Ritter Sport* rejected the accusation. They went to court and won the trial. *Stiftung Warentest* had to withdraw the issue in question of *Test* magazine. This affair was all over the media in January of 2014. *Stiftung Warentest* and their quality label, also called *Test*, made negative headlines.

HILDENBRAND AND KÜHL (2014) argue that *Ritter Sport*'s court victory over *Stiftung Warentest* is a Pyrrhic victory for both high-quality producers and quality-conscious consumers. Their argumentation is based on empirical evidence that the credibility of a quality label stems, inter alia, from the credibility of the awarding organization and that the credibility of the awarding organization stems, inter alia, from the absence of negative headlines (see DR. GRIEGER & CIE. MARKTFORSCHUNG 2013). From the presence of negative headlines on *Stiftung Warentest*, HILDENBRAND AND KÜHL (2014) reason that the credibility of the *Test* label is undermined. Only if this is true, ceteris paribus, information is lost and the court victory is a Pyrrhic victory.

Using the *Ritter Sport* versus *Stiftung Warentest* case, we analyze whether negative headlines really undermine the credibility of a quality label. In addition, we examine what can be done to restore or, more generally, increase the credibility of a quality label. Based on a quasi-natural experiment (see DiNARDO 2008), we find that the negative headlines on *Stiftung*

*Warentest* have undermined the credibility of the *Test* label. We also find that the credibility of the *Test* label can be restored by providing reference values to the tests, strengthening the independence of *Stiftung Warentest*, and using laboratory methods in the tests. These findings are in line with general findings.

The paper is organized as follows. In the next section, the hypotheses are deduced and presented. The experimental design and the procedures are introduced in the section after next. After that, the experimental results are summarized and discussed. We conclude in the last section.

## 2 Hypotheses

In general, the (perceived) credibility of a source of communication (source credibility) stems from three dimensions (see EISEND 2006a, 2006b): the source is expected to tell the truth (the inclination toward truth), the source is expected to know the truth (the potential of truth), and the presentation. The presentation dimension covers visible characteristics like the attributes of a source (see HALEY 1996; JAVALGI ET AL. 1994; SCHUMANN, HATHCOTE, AND WEST 1991) or the information on a source (see KLEBBA AND UNGER 1983). Negative information is regularly found to decrease source credibility, and positive information is usually found to increase source credibility (see KLEBBA AND UNGER 1983). The more credible a source is, the more persuasive it will be (see PORNPITAKPAN 2004; GIERL, STICH, AND STROHMAYR 1997; STERNTHAL, PHILLIPS, AND DHOLAKIA 1978).

### *Negative Headlines Undermine the Credibility of a Quality Label*

The concept of source credibility can be applied to organizations awarding quality labels like *Stiftung Warentest*. The more credible an awarding organization is, the more persuasive or, more precisely, informative a quality label will be (signal credibility). Only if signal credibility is given, a quality label can serve as quality indicator. Because foods are typically not search goods but experience or credence goods, consumers cannot check the quality before their purchases (see NELSON 1970; DARBY AND KARNI 1973). For example, take a bar of hazelnut chocolate. It has several attributes: search attributes like the price, experience attributes like the taste, and credence attributes like the origin of the ingredients. Both the experience attribute and the credence attribute can be turned into search attributes by a quality label. They are turned into search attributes if it is credible.

To put it in different words, if a quality label is credible, producers no longer possess more information about invisible characteristics than consumers. There will be no information asymmetry any longer (see MOUSSA AND TOUZANI 2008, p. 527). If the credibility of a quality label is undermined, *ceteris paribus*, the willingness to pay will decrease. Depending on the extent of the decrease, high-quality products may be driven out of the market by low-quality products. That is, adverse selection may arise (see AKERLOF 1970). High-quality producers and quality-conscious consumers would be the victims of the information loss. Of course, other quality indicators exist. For example, advertising or warranties are discussed (see NELSON 1974; GROSSMAN 1981).

*Stiftung Warentest* classifies products on the basis of five grades: very good, good, satisfactory, adequate, and unsatisfactory. If a product is classified as very good or good, the *Test* label indicates high quality. In this instance, it can serve as a quality label, and producers normally print the *Test* label on the packaging of their products. More than 90 percent of the German consumers know the *Test* label (see DR. GRIEGER & CIE. MARKTFORSCHUNG 2013, p. 12; EPP ET AL. 2010, p. 61; VERBRAUCHERZENTRALE BUNDESVERBAND 2008, p. 18), and more than 75 percent of the German consumers knowing the *Test* label trust in its credibility (see DR. GRIEGER & CIE. MARKTFORSCHUNG 2013, p. 12–13; Nestlé 2012, p. 22). In general, the consumer protection activities of *Stiftung Warentest* are regarded as being the most effective ones (see VERBRAUCHERZENTRALE BUNDESVERBAND 2008, p. 14).

The German consumers trust in the credibility of a quality label if (descending order) the awarding organization is independent (1), reference values are given (2), the methodology is transparent (3), laboratory methods are used (4), and the absence of negative headlines (5) (see DR. GRIEGER & CIE. MARKTFORSCHUNG 2013, p. 18–19). For about 60 percent of the German consumers, the absence of negative headlines is essential for trusting in the credibility of a quality label (see DR. GRIEGER & CIE. MARKTFORSCHUNG 2013, p. 18–19).

Hypothesis 1: The negative headlines on *Stiftung Warentest* represent negative information in the sense of the theory of source credibility. They have undermined the credibility of the *Test* label.

### *Independence, Reference Values, Transparency, and Laboratory Methods Increase the Credibility of a Quality Label*

Because negative information is usually found to have a greater impact than positive information (see KLEBBA AND UNGER 1983), especially the other attributes have to be addressed in order to increase the credibility of a source. For about 70 percent of the German consumers, the independence of an awarding organization is essential for trusting in the credibility of a quality label. The same holds true for reference values. For about 60 percent of the German consumers, methodological transparency is essential. The same holds true for laboratory methods (see DR. GRIEGER & CIE. MARKTFORSCHUNG 2013, p. 18–19).

From a theoretical point of view, these attributes are present to a large extent. However, there are weaknesses.

*Stiftung Warentest* (2011) seems to be independent (1) because they are a foundation under civil law. The endowment capital is 75 million euros. It has been given by the Federal Republic of Germany. However, the independence is reduced by a license fee. It has to be paid by producers that want to print the *Test* label on the packaging of their products. Because the *Test* label can only serve as a quality label if a product is classified as very good or good, an incentive for grade inflation exists. Because the license fee is more expensive if television and cinema advertising is included, an incentive for grade distortion exists if big producers do more television and cinema advertising than small producers. This seems to hold true.

*Stiftung Warentest* (2013) gives reference values (2). Besides the *Test* logo, the *Test* label consists of a quality grade. The number of products in the test is also given, and there is a reference to the issue in question of *Test* magazine. However, the choice of products is unknown. The choice is said to be “on the basis of market research and in accordance with the specified test criteria,” but details are unknown (*Stiftung Warentest* 2014a, p. 2). The methodology seems to be transparent (3), and *Stiftung Warentest* (2014b) uses laboratory methods (4). However, in the *Ritter Sport* versus *Stiftung Warentest* case, the court complained that *Stiftung Warentest* did not reveal the conditions of their test and that their interpretation of the regulation on flavorings was wrong and misleading (RUHWINKEL 2014).

Hypothesis 2.1: More (perceived) independence can increase the credibility of the *Test* label.

Hypothesis 2.2: The (perceived) presence of reference values is positively related to the credibility of the *Test* label. Hypothesis 2.3: More (perceived) transparency can increase the credibility of the *Test* label.

Hypothesis 2.4: The (perceived) presence of laboratory methods is positively related to the credibility of the *Test* label.

## **3 Experimental Design and Procedures**

### **3.1 Experimental Design**

There are four treatments: BASELINE, INFORMATION, RECALL, and AMPLIFICATION. An overview of the treatments is given in Table 1. In BASELINE and RECALL, no additional information is given. BASELINE serves as a control group. If a participant does

not remember any headlines or reporting on the Test label or Stiftung Warentest, s/he is assigned to BASELINE. Otherwise, s/he is assigned to RECALL. Because BASELINE and RECALL naturally occur, those are natural treatments.

**Table 1: Treatments**

		<b>Random matching!</b>	
<b>natural variation</b>	<b>artificial variation</b>	<b>no headlines are shown</b>	<b>negative headlines are shown</b>
<b>Do you remember any headlines?</b>	<b>no, I do not</b>	BASELINE	INFORMATION
	<b>yes, namely negative headlines</b>	RECALL	AMPLIFICATION

INFORMATION and AMPLIFICATION are artificial treatments because additional information is given. Regardless of whether a participant remembers or does not remember any headlines or reporting, negative headlines are shown. Whether additional information is shown or not shown is determined by random matching:  $P(\text{BASELINE} \cup \text{RECALL}) = \frac{1}{2}$  and  $P(\text{INFORMATION} \cup \text{AMPLIFICATION}) = 1 - P(\text{BASELINE} \cup \text{RECALL}) = \frac{1}{2}$ .

### 3.2 Procedures

The quasi-natural experiment was conducted at Justus Liebig University Giessen in May and June of 2014. All students were invited to participate in a survey. A link to the survey was electronically mailed to them. The survey was posted on the website of the university on May 22, 2014. A reminder was sent on June 16, 2014. The survey return deadline was July 3, 2014. (The electronic questionnaire can be obtained upon request.)

In order to motivate the students to join the survey, we conducted a raffle. There were ten prizes in our raffle. Each prize consisted of 10 euros. The winners were randomly drawn from the sample of completed questionnaires. They were informed via electronic mail. Seven prizes were collected. Three winners did not show up (retrieved February 18, 2015.) Overall, 542 students participated. However, 169 questionnaires were not completed. These questionnaires were excluded. Therefore, we were left with 373 participants.

## 4 Experimental Results

In total, 276 participants (73.99 percent) were female; 97 participants (26.01 percent) were male. On average, a participant was 24.64 years old with a standard deviation of 4.79 years. The youngest participant was 16 years old. The oldest one was 50 years old. Most participants belonged to the Department of Agricultural Sciences, Nutritional Sciences, and Environmental Management.

Quality labels are important in general, and quality labels are even more important when foods are bought. To elicit the importance of quality labels, a slide switch was given to the participants. Using the slide switch, the participants could set a value between 0 indicating no importance and 100 indicating high importance.

On average, the participants set a value of 57.26 with a standard deviation of 22.20 in general and a value of 63.16 with a standard deviation of 24.43 when foods are bought. The difference is significant (one-sided paired  $t$ -test:  $p = 0.000$ ). If the values between 51 and 100 are regarded as indicating importance, quality labels are important for 66.22 percent of the participants in general and for 74.80 percent of the participants when food are bought. The

difference is also significant (one-sided paired  $t$ -test:  $p = 0.000$ ). This is in line with the findings for German consumers (DR. GRIEGER & CIE. MARKTFORSCHUNG 2013, p. 10).

The participants trust in the credibility of a quality label if (descending order) the methodology is transparent (1), the awarding organization is independent (2), it is up to date (3), laboratory methods are used (4), there are no negative headlines (5), reference values are given (6), it is present on the packaging of many products (7), it is advertised (8), and it is used for advertising (9). To elicit the attributes for trusting in the credibility of a quality label, the above-mentioned slide switch was given to the participants. This is quite similar to the findings for German consumers (DR. GRIEGER & CIE. MARKTFORSCHUNG 2013, p. 18).

For 81.77 percent of the participants, the absence of negative headlines is essential for trusting in the credibility of a quality label if the values between 51 and 100 are regarded as indicating importance. On average, the participants set a value of 67.85 with a standard deviation of 22.19. This is also similar to the findings for German consumers (DR. GRIEGER & CIE. MARKTFORSCHUNG 2013, p. 19).

There was one participant who did not know the *Test* label. This participant was filtered out, and 372 participants or 99.73 percent of the participants knew the *Test* label. This is slightly more than in the representative sample used by DR. GRIEGER & CIE. MARKTFORSCHUNG (2013, p. 12). It may be caused by self-selection.

An overview of the number of participants in the treatments is given in Table 2. Most participants were assigned to BASELINE. The fewest participants were assigned to AMPLIFICATION.

**Table 2: Treatments and participants**

treatments	remembrance	headlines	participants	females	males
BASELINE	no	no	150	115	35
INFORMATION	no	yes	130	36	94
RECALL	yes	no	50	16	34
AMPLIFICATION	yes	yes	42	10	32

We do not have a representative sample neither of the general population in Germany, nor of the student population at Justus Liebig University Giessen. That is why we have to carefully interpret our results in the following sections. However, treatment effects can be fully interpreted because a representative sample is not needed here. This is the reason for our experimental design. That is why we focus on treatment effects.

#### *Do Negative Headlines Undermine the Credibility of a Quality Label?*

Most participants trust in the credibility of the *Test* label. To elicit its credibility, a slide switch was given to the participants again. Using the slide switch, the participants could set a value between 0 indicating no credibility and 100 indicating high credibility.

On average, the participants set a value of 68.26 with a standard deviation of 20.54. If the values between 51 and 100 are regarded as indicating trust in the credibility of the *Test* label, more than 83.33 percent of the participants knowing the *Test* label trust in its credibility. In comparison, more than 75 percent of the German consumers knowing the *Test* label trust in its credibility (Dr. Grieger & Cie. Marktforschung 2013, p. 13).

Whether the negative headlines on *Stiftung Warentest* have undermined the credibility of the *Test* label can be analyzed by comparing BASELINE to RECALL or INFORMATION. By comparing BASELINE to RECALL, the natural treatments are analyzed. The advantage is that we compare a situation of cold cognition to a situation of cold cognition. In both treatments, no additional information regarding the *Test* label is given to the participants.

Therefore, no new information has to be processed. The disadvantage is that there is no random matching but self-selection. A self-selection bias may be present.

By comparing BASELINE to INFORMATION, there is a random matching. This is an advantage. However, we compare a situation of cold cognition to a situation of hot cognition. In INFORMATION, additional information regarding the *Test* label is given to the participants. Therefore, new information has to be processed. Because no new information has to be processed in BASELINE, we have an asymmetric comparison. This is a disadvantage. That is why we make both comparisons.

On average, the participants set a value of 71.25 with a standard deviation of 18.40 in BASELINE. In RECALL, they set a mean value of 67.14 with a standard deviation of 24.45. However, 31 participants recall positive headlines. If these participants are excluded, the participants in RECALL set a mean value of 52.37 with a standard deviation of 28.49. The difference is significant (one-sided unpaired  $t$ -test:  $p = 0.000$  with equal variances and  $p = 0.005$  with unequal variances; one-sided Mann-Whitney  $U$ -test because of the small sample size:  $p = 0.003$ ). If the values between 51 and 100 are regarded as indicating trust in the credibility of the *Test* label, 87.33 percent of the participants in BASELINE and 52.63 percent of the participants in RECALL trust in its credibility. The difference is also significant (one-sided unpaired  $t$ -test:  $p = 0.000$  with equal variances and  $p = 0.005$  with unequal variances; one-sided Mann-Whitney  $U$ -test because of the small sample size:  $p = 0.000$ ).

In INFORMATION, the participants set a mean value of 67.62 with a standard deviation of 19.84. The difference is significant (one-sided unpaired  $t$ -test:  $p = 0.057$  with equal variances and  $p = 0.058$  with unequal variances; one-sided Mann-Whitney  $U$ -test as a supplement:  $p = 0.099$ ). If the values between 51 and 100 are regarded as indicating trust in the credibility of the *Test* label, 83.85 percent of the participants in INFORMATION trust in its credibility. The difference is not significant (one-sided unpaired  $t$ -test:  $p = 0.204$  with equal variances and  $p = 0.205$  with unequal variances; one-sided Mann-Whitney  $U$ -test because of the small sample size:  $p = 0.203$ ).

Therefore, negative headlines can undermine the credibility of a quality label. Hypothesis 1 is supported. The negative headlines on *Stiftung Warentest* represent negative information in the sense of theory of source credibility. They have undermined the credibility of the *Test* label. The undermining of its credibility is more excessive in RECALL than in INFORMATION (one-sided unpaired  $t$ -test:  $p = 0.002$  with equal variances and  $p = 0.018$  with unequal variances; one-sided Mann-Whitney  $U$ -test because of the small sample size:  $p = 0.012$  if the values between 0 and 100 are analyzed | one-sided unpaired  $t$ -test:  $p = 0.001$  with equal variances and  $p = 0.009$  with unequal variances; one-sided Mann-Whitney  $U$ -test because of the small sample size:  $p = 0.002$  if the values between 51 and 100 are regarded as indicating trust in the credibility of the *Test* label). The impact of cold cognition is stronger than the impact of hot cognition.

What if the participants who recall negative headlines are shown negative headlines? In AMPLIFICATION, the participants set a mean value of 60.93 with a standard deviation of 23.30. However, 19 participants recall positive headlines. If these participants are excluded, the participants in AMPLIFICATION set a mean value of 53.83 with a standard deviation of 25.33. The difference between RECALL and AMPLIFICATION is not significant (two-sided unpaired  $t$ -test:  $p = 0.862$  with equal variances and  $p = 0.863$  with unequal variances; two-sided Mann-Whitney  $U$ -test because of the small sample size:  $p = 0.940$ ). The difference is also not significant (two-sided unpaired  $t$ -test:  $p = 0.421$  with equal variances and  $p = 0.423$  with unequal variances; two-sided Mann-Whitney  $U$ -test because of the small sample size:  $p = 0.413$ ) if the values between 51 and 100 are regarded as indicating trust in the credibility of the *Test* label. Therefore, the undermining of its credibility is equally excessive. The observations of RECALL and AMPLIFICATION can be pooled for further analyses. The differences between the pooled observations and BASELINE (one-sided unpaired  $t$ -test:  $p =$



0.000 with equal variances and  $p = 0.000$  with unequal variances; one-sided Mann-Whitney  $U$ -test as a supplement:  $p = 0.000$  if the values between 0 and 100 are analyzed | one-sided unpaired  $t$ -test:  $p = 0.000$  with equal variances and  $p = 0.001$  with unequal variances; one-sided Mann-Whitney  $U$ -test as a supplement:  $p = 0.000$  if the values between 51 and 100 are regarded as indicating trust in the credibility of the *Test* label) or INFORMATION (one-sided unpaired  $t$ -test:  $p = 0.000$  with equal variances and  $p = 0.001$  with unequal variances; one-sided Mann-Whitney  $U$ -test as a supplement:  $p = 0.001$  if the values between 0 and 100 are analyzed | one-sided unpaired  $t$ -test:  $p = 0.000$  with equal variances and  $p = 0.003$  with unequal variances; one-sided Mann-Whitney  $U$ -test as a supplement:  $p = 0.001$  if the values between 51 and 100 are regarded as indicating trust in the credibility of the *Test* label) remain significant.

This result is bad news for *Stiftung Warentest*. Consumers do not forgive *Stiftung Warentest* just because of having processed the negative headlines. For *Stiftung Warentest*, the only consolation is that many participants do not recall the negative headlines. However, because our sample is not representative, this number cannot be fully interpreted.

### *Do Independence, Reference Values, Transparency, and Laboratory Methods Increase the Credibility of a Quality Label?*

Whether independence, reference values, transparency, and laboratory methods increase the credibility of the *Test* label can be analyzed by a regression analysis. An econometric model to explain the credibility of the *Test* label (*credibility*) in terms of the independence of *Stiftung Warentest* (*independence*), the presence of reference values (*reference*), the transparency of the methodology (*transparency*), and the use of laboratory methods (*laboratory*) is

$$credibility = \beta_0 + \beta_1 independence + \beta_2 reference + \beta_3 transparency + \beta_4 laboratory + u.$$

To elicit the regressors, a slide switch was given to the participants again. Using the slide switch, the participants could set a value between 0 indicating no independence/reference values/transparency/laboratory methods and 100 indicating the opposite. Hence, both the regressand and the regressors can take values between 0 and 100. Descriptions and summary statistics of the variables are given in Table 3.

**Table 3: Variables, descriptions and summary statistics**

variable	description	mean value	standard deviation	minimum	maximum
<i>credibility</i>	credibility of the <i>Test</i> label	68.26	20.54	0	100
<i>independence</i>	independence of <i>Stiftung Warentest</i>	54.78	24.16	0	100
<i>reference</i>	presence of reference values	61.35	23.15	0	100
<i>transparency</i>	transparency of the methodology	45.30	25.01	0	100
<i>laboratory</i>	use of laboratory methods	64.20	21.33	0	100

We estimate the model from above by ordinary least squares (model 1). Because heteroskedasticity is present, White robust standard errors are used. The residuals are approximately normally distributed. There is no multicollinearity, and the linearity assumption holds. We also estimate an extended model with additional regressors: INFORMATION, RECALL, and AMPLIFICATION are included in the model as dummy

variables (model 2). The estimated values are given in Table 4. Standard errors and  $p$ -values for the standard  $t$ -tests are shown in parentheses.

**Table 4: Estimated values for both models**

regressor	model 1	model 2
<i>constant</i>	20.53922 (3.062006, $p = 0.000$ )	21.51579 (3.087169, $p = 0.000$ )
<i>independence</i>	0.1969372 (0.0464305, $p = 0.000$ )	0.1903446 (0.0462929, $p = 0.000$ )
<i>reference</i>	0.4059006 (0.0539327, $p = 0.000$ )	0.4058051 (0.0533845, $p = 0.000$ )
<i>transparency</i>	-0.0117563 (0.039833, $p = 0.768$ )	-0.0140468 (0.0398484, $p = 0.725$ )
<i>laboratory</i>	0.195699 (0.0543817, $p = 0.000$ )	0.2052884 (0.0531927, $p = 0.000$ )
<i>information</i>	-	0.3224762 (1.584524, $p = 0.839$ )
<i>recall</i>	-	-2.177117 (2.288876, $p = 0.342$ )
<i>amplification</i>	-	-8.339684 (2.819158, $p = 0.003$ )
number of observations	372	372
coefficient of determination	0.5464	0.5636

There is not much difference between the models. In both models, the transparency of the methodology has no significant effect on the credibility of the *Test* label. The other estimated coefficients are significant and positive. That is, independence, reference values, and laboratory methods determine the credibility of the *Test* label. Hypotheses 2.1, 2.2, and 2.4 are supported. More (perceived) independence can increase the credibility of the *Test* label. The (perceived) presence of reference values and laboratory methods is positively related to the credibility of the *Test* label. Hypothesis 2.3 is not supported. More (perceived) transparency cannot increase the credibility of the *Test* label.

## 5 Discussion

The credibility of a quality label depends, inter alia, on the independence of the awarding organization and the methodological transparency. As mentioned above, the independence of *Stiftung Warentest* can be questioned because it is partially funded by license fees. Buying the usage rights for the *Test* label is only attractive for products with satisfying test results. This creates an incentive for grade inflation. In addition, *Stiftung Warentest* charges higher fees for the right to use the *Test* label for television or cinema advertising. These forms of advertising are primarily expected to be demanded by producers with comparatively high advertising budgets. This creates an incentive not only for grade distortion, but also for selective selection. That is, the choice of the products for the tests may be more influenced by the advertising budgets of the producers than by the preferences of the consumers.

The use of laboratory methods and the transparency of the methodology are quite similar. Methodological transparency can be seen as a generalization. If it is known that laboratory methods are used in a test, there is methodological transparency regarding this. What does that mean in detail? Regarding the *Test* label, methodological transparency is questionable because the choice of the products in the tests is not revealed. Because of missing indications

regarding the test criteria and other products in the test, the *Test* label actually does neither provide data nor information on that at the time of the purchase decision. Consumers do not only not know the underlying criteria. They do also not know whether the test criteria and weights of the criteria correspond with their preferences. In fact, this involves some kind of paternalism, which has not been questioned or objected so far: neither by consumers nor by politicians.

Information that is also not available at the point of sale is the reason why a label-free competing product is label-free. There are two reasons that can be possible: either the competing product was not tested for an unknown reason or its producer was not willing to buy the license because of a bad test result or (just) a shortage of financial resources.

However, this problem of information could be easily solved by the provision of additional information at the point of sale. For example, the *Test* label could be extended by a QR code. In addition, a QR code scanner could be offered as a charged mobile app. With a modus operandi like that, the spirit of the time would be hit. New target groups would be addressed. The revenues, generated by this app, could substitute current earnings from the license fee.

Finding new target groups is necessary because the circulation of *Test* magazine is declining. The paid circulation of *Test* magazine almost halved since 1991. Back then, 960.000 copies were sold. In 2013, only 455.000 copies could be sold (see *Stiftung Warentest* 2014c). In comparison, the demand for the content on the website of *Stiftung Warentest* continuously increases (see *Stiftung Warentest* 2014d).

In 2014, approximately 40 million Germans own a smartphone (see comScore 2014). About half of them use their smartphone to access the internet several times per day (see Tomorrow Focus AG 2014). According to projections, the spread of smartphones will increase worldwide (see eMarketer 2014). Up to half of the smartphone users already scanned QR codes for further information (see MGH 2011; Nielsen 2012, p. 4). Hence, a real chance exists to generate new earnings.

## 6 Conclusion

We find quality labels are important in general, and quality labels are even more important when foods are bought. This is in line with the findings for German consumers. The participants trust in the credibility of a quality label if (descending order) the methodology is transparent, the awarding organization is independent, it is up to date, laboratory methods are used, there are no negative headlines, reference values are given, it is present on the packaging of many products, it is advertised, and it is used for advertising. This is also in line with the findings for German consumers. For 81.77 percent of the participants, the absence of negative headlines is essential for trusting in the credibility of a quality label. This is also quite similar to the findings for German consumers.

Using the *Ritter Sport* versus *Stiftung Warentest* case, we analyzed whether negative headlines really undermined the credibility of a quality label. In addition, we examined what could be done to restore or, more generally, increase the credibility of a quality label. Based on a quasi-natural experiment, we find that the negative headlines on *Stiftung Warentest* have undermined the credibility of the *Test* label. The negative headlines on *Stiftung Warentest* represent negative information in the sense of theory of source credibility. Therefore, information is lost and the court victory is a Pyrrhic victory for both high-quality producers and quality-conscious consumers.

Of course, this result is bad news for *Stiftung Warentest*. The sales of *Test* magazine may be negatively affected. Because the undermining of the credibility of the *Test* label is most excessive if negative headlines are remembered, consumers being interested in high-quality products are especially affected. These consumers typically subscribe to relevant magazines like *Test* magazine. They may remember the negative headlines very well. As a consequence, they may cancel their subscriptions. Whether the sales or subscriptions of *Test* magazine are

affected in reality is not known at the moment. What is known is that the sales of *Test* magazine almost halved since 1991. Hence, *Stiftung Warentest* cannot afford to do nothing. They have to do something. Increasing the credibility of the *Test* label is one possibility. Advertising the *Test* label is another.

We find that the credibility of the *Test* label can be increased by providing reference values to the tests, strengthening the independence of *Stiftung Warentest*, and using laboratory methods in the tests. These findings are in line with general findings.

In addition to the presented results, we focused on arbitrary quality labels. Most of our findings are valid for them as well. Moreover, we identified two main sources of the credibility of a quality label by conducting a principal component analysis: hard facts concerning the awarding organization/the quality label and the presence of the quality label. These sources provide scope for further conclusions. However, the details cannot be presented here for the lack of space.

## References

- AKERLOF, G. A. (1970), "The Market for 'Lemons': Quality Uncertainty and the Market Mechanism," *Quarterly Journal of Economics*, 84 (3), 488–500.
- COMSCORE (2014), "Anzahl der Smartphone-Nutzer in Deutschland in den Jahren 2009 bis 2014 (in Millionen)," Statista (accessed January 4, 2015), [available at <http://de.statista.com/statistik/daten/studie/198959/umfrage/anzahl-der-smartphonenuutzer-in-deutschland-seit-2010>].
- DARBY, M. R. AND E. KARNI (1973): "Free Competition and the Optimal Amount of Fraud," *Journal of Law and Economics*, 16 (1), 67–88.
- DINARDO, J. (2008), "Natural Experiments and Quasi-Natural Experiments," in *New Palgrave Dictionary of Economics*, Steven N. Durlauf and Lawrence E. Blume, eds. London: Palgrave Macmillan, 856–859.
- DR. GRIEGER & CIE. MARKTFORSCHUNG (2013), "Gütesiegel in Deutschland: Repräsentative Bevölkerungsbefragung zu Bekanntheit, Relevanz und Vertrauen bei Gütesiegeln, unter Spezialbetrachtung von Servicesiegeln," research report, Dr. Grieger & Cie. Marktforschung, Hamburg.
- EISEND, M. (2006a), "Source Credibility Dimensions in Marketing Communication: A Generalized Solution," *Journal of Empirical Generalisations in Marketing Science*, 10 (2), 1–33.
- EISEND, M. (2006b), "Source Credibility in Marketing Communication: A Meta-Analysis," *Marketing: Journal of Research and Management*, 2 (1), 43–60.
- EMARKETER (2014), "Prognose zum Anteil der Smartphone-Nutzer weltweit von 2013 bis 2018," Statista (accessed January 4, 2015), [available at <http://de.statista.com/statistik/daten/studie/321204/umfrage/prognose-zum-anteil-der-smartphone-nutzer-weltweit>].
- EPP, A.; B. MICHALSKI AND U. BANASIAK (2010), *Pflanzenschutzmittel-Rückstände in Lebensmitteln: Die Wahrnehmung der deutschen Bevölkerung – Ein Ergebnisbericht*. Berlin: BfR Wissenschaft.
- GIERL, H.; A. STICH AND M. STROHMAYR (1997), "Einfluß der Glaubwürdigkeit einer Informationsquelle auf die Glaubwürdigkeit der Information," *Marketing: Zeitschrift für Forschung und Praxis*, 19 (1), 27–31.
- GROSSMAN, S. J. (1981), "The Informational Role of Warranties and Private Disclosure about Product Quality," *Journal of Law and Economics*, 24 (3), 461–483.
- HALEY, E. (1996), "Exploring the Construct of Organization as Source: Consumers' Understandings of Organizational Sponsorship of Advocacy Advertising," *Journal of Advertising*, 25 (2), 19–35.

- HILDENBRAND, A. AND R. KÜHL (2014), "Ritter Sport und Stiftung Warentest: Informationsdefizite überwinden," *Wirtschaftsdienst*, 94 (3), 217–220.
- JAVALGI, R. G., M. B. TRAYLOR, A. C. GROSS, AND E. LAMPMAN (1994), "Awareness of Sponsorship and Corporate Image: An Empirical Investigation," *Journal of Advertising*, 23 (4), 47–58.
- KLEBBA, J. M. AND L. S. UNGER (1983), "The Impact of Negative and Positive Information on Source Credibility in a Field Setting," *Advances in Consumer Research*, 10 (1), 11–16.
- MGH (2011), "QR Code Usage and Interest Survey", MGH (accessed January 4, 2015), [available at <http://mghus.com/files/6213/5292/5743/QRcodeStats3.23.11.pdf>].
- MOUSSA, S. AND M. TOUZANI (2008), "The Perceived Credibility of Quality Labels: A Scale Validation with Refinement," *International Journal of Consumer Studies*, 32 (5), 526–533.
- NELSON, P. (1970), "Information and Consumer Behavior," *Journal of Political Economy*, 78 (2), 311–329.
- NELSON, P. (1974), "Advertising as Information," *Journal of Political Economy*, 82 (4), 729–754.
- NESTLÉ (2012), *Das is(s)t Qualität: Auszüge aus der Nestlé-Studie 2012*. Frankfurt: Nestlé Deutschland.
- NIELSEN (2012), "Bei welchen Gelegenheiten haben Sie bisher QR-Codes gescannt?," Statista (accessed January 4, 2015), [available at <http://de.statista.com/statistik/daten/studie/243059/umfrage/umfrage-zu-den-gruenden-der-bisherigen-nutzung-von-qr-codes/>].
- PORNPITAKPAN, C. (2004), "The Persuasiveness of Source Credibility: A Critical Review of Five Decades' Evidence," *Journal of Applied Social Psychology*, 34 (2), 243–281.
- RUHWINKEL, S. (2014), "Streit ums Schokoladen-Aroma: Gericht verbietet Äußerungen eines Warentestinstituts," press release, Landgericht I, München.
- SCHUMANN, D. W., J. M. HATHCOTE, AND S. WEST (1991), "Corporate Advertising in America: A Review of Published Studies on Use, Measurement, and Effectiveness," *Journal of Advertising*, 20 (3), 35–56.
- STERNTHAL, B., L. W. PHILLIPS, AND R. DHOLAKIA (1978), "The Persuasive Effect of Source Credibility: A Situational Analysis," *Public Opinion Quarterly*, 42 (3), 285–314.
- STIFTUNG WARENTEST (2011), *Satzung*. Berlin: Stiftung Warentest.
- STIFTUNG WARENTEST (2013), *Licensing Agreement*. Berlin: Stiftung Warentest.
- STIFTUNG WARENTEST (2014a), *Objective and Independent*. Berlin: Stiftung Warentest.
- STIFTUNG WARENTEST (2014b), "Testablauf," Stiftung Warentest (accessed May 13, 2014), [available at <http://www.test.de/unternehmen/testablauf>].
- STIFTUNG WARENTEST (2014c), "Verkaufsentwicklung," Stiftung Warentest (accessed January 5, 2015), [available at <https://www.test.de/filestore/auflage-test-gross.jpg?path=/11/a7/18ecaab9-1a02-4997-be01-a55372005ca1-file.jpg&key=88D0202AB0DF7A82C9BE92F2ABBC5C316582A0119>].
- STIFTUNG WARENTEST (2014d), "test.de," Stiftung Warentest (accessed January 5, 2015), [available at <https://www.test.de/filestore/testde-gross.jpg?path=/91/a0/902f2895-b38b-44ea-a728-32b00ff512a1-file.jpg&key=FEC02405A07AB3FE0D417B189EB100501EA4B1C7>].
- TOMORROW FOCUS AG (2014), "Wie oft nutzen Sie mobiles Internet mit Ihrem Mobiltelefon?," Statista (accessed January 4, 2015), [available at <http://de.statista.com/statistik/daten/studie/167623/umfrage/nutzungshaeufigkeit-des-mobilen-internets-ueber-das-mobiltelefon>].
- VERBRAUCHERZENTRALE BUNDESVERBAND (2008), *Verbraucherschutz in Deutschland: Was meinen die Verbraucher?* Berlin: Verbraucherzentrale Bundesverband.