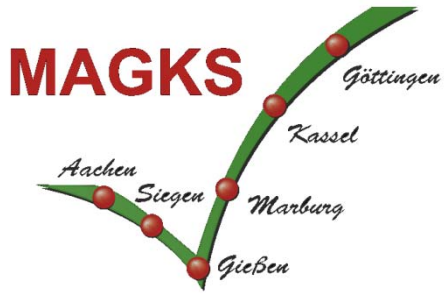

Joint Discussion Paper Series in Economics

by the Universities of

Aachen · Gießen · Göttingen
Kassel · Marburg · Siegen

ISSN 1867-3678

No. 14-2015

Andreas Hildenbrand, Rainer Kühl and Anne Piper

Do negative headlines really undermine the credibility of a quality label? A quasi-natural experiment

This paper can be downloaded from
http://www.uni-marburg.de/fb02/makro/forschung/magkspapers/index_html%28magks%29

Coordination: Bernd Hayo • Philipps-University Marburg
School of Business and Economics • Universitätsstraße 24, D-35032 Marburg
Tel: +49-6421-2823091, Fax: +49-6421-2823088, e-mail: hayo@wiwi.uni-marburg.de

Do Negative Headlines Really Undermine the Credibility of a Quality Label?

A Quasi-Natural Experiment

1 June 2015

Andreas Hildenbrand, Dr. rer. pol. (corresponding author), Institute for Agribusiness Management and Food Economics, Justus Liebig University Giessen, Senckenbergstrasse 3, 35390 Giessen, Germany, +49 641 99-37272, andreas.hildenbrand@agrar.uni-giessen.de.

Rainer Kühl, Prof. Dr. sc. agr., Institute for Agribusiness Management and Food Economics, Justus Liebig University Giessen, Senckenbergstrasse 3, 35390 Giessen, Germany, +49 641 99-37270, rainer.kuehl@agrar.uni-giessen.de.

Anne Piper, M. Sc., Institute for Agribusiness Management and Food Economics, Justus Liebig University Giessen, Senckenbergstrasse 3, 35390 Giessen, Germany, +49 641 99-37275, anne.piper@agrar.uni-giessen.de.

Acknowledgements

For useful suggestions, we thank participants at the 2014 GfeW Annual Conference.

Do Negative Headlines Really Undermine the Credibility of a Quality Label?

A Quasi-Natural Experiment

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 labelling an artificial flavouring as a natural flavouring. *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. Using the *Ritter Sport* versus *Stiftung Warentest* case, we analyse whether negative headlines really undermine the credibility of a quality label by examining *Stiftung Warentest* and their quality label, also called *Test*. 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 increased by providing reference values to the tests, strengthening the independence of *Stiftung Warentest*, and using laboratory methods in the tests. For the most part, the same holds true for any quality label.

Keywords

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

Executive Summary (German)

Im Jahr 2013 testete Stiftung Warentest Nuss-Schokoladen für das Dezember-Heft ihrer Test-Zeitschrift. Stiftung Warentest ist eine der wichtigsten deutschen Verbraucherorganisationen. Ritter Sport ist ein deutscher Schokoladenhersteller, der Schokoladen von guter Qualität produziert. Im Test fiel die Voll-Nuss-Schokolade von Ritter Sport durch: Note „mangelhaft". Begründet wurde die schlechte Note nicht mit schlechter Qualität, sondern mit einer irreführenden Etikettierung. Der Vorwurf lautete, Ritter Sport hätte ein künstliches Aroma als natürliches Aroma bezeichnet. Ritter Sport bestritt dies, ging vor Gericht und gewann den Prozess. Infolgedessen musste Stiftung Warentest das Test-Heft zurückziehen. Dieser Vorfall war im Januar 2014 in allen überregionalen Medien präsent. Stiftung Warentest und ihr Qualitätssigel, das Testsiegel, machten Negativschlagzeilen.

Hildenbrand/Kühl (2014) argumentieren, dass der Sieg von Ritter Sport über Stiftung Warentest ein Pyrrhussieg ist: sowohl für die Hersteller, die gute Qualität produzieren, als auch für die Endverbraucher, die ein Qualitätsbewusstsein haben. Ihrer Argumentation liegt zugrunde, dass die Glaubwürdigkeit eines Qualitätssigels von der Glaubwürdigkeit des Emittenten abhängt, und dass die Glaubwürdigkeit eines Emittenten (unter anderem) von der Abwesenheit von Negativschlagzeilen herrührt (vgl. *Dr. Grieger & Cie. Marktforschung* 2013). Aus der Anwesenheit von Negativschlagzeilen über Stiftung Warentest schließen Hildenbrand und Kühl, dass die Glaubwürdigkeit des Testsiegels untergraben worden ist. Nur wenn das stimmt, hat sich ein Informationsverlust ergeben, der Produzenten guter Qualität und Endverbraucher mit Qualitätsbewusstsein schlechterstellt, das heißt, ist ein Pyrrhussieg erstritten worden.

Anhand des Ritter-Sport-Stiftung-Warentest-Rechtsstreits analysieren wir, ob Negativschlagzeilen die Glaubwürdigkeit eines Qualitätssiegels tatsächlich untergraben. Zudem untersuchen wir, was getan werden kann, um die Glaubwürdigkeit eines Qualitätssiegels wiederherzustellen oder, allgemeiner ausgedrückt, zu erhöhen. Mithilfe eines quasi-natürlichen Experiments zeigen wir, dass die Negativschlagzeilen über Stiftung Warentest die Glaubwürdigkeit des Testsiegels untergraben haben. Die Negativschlagzeilen über Stiftung Warentest stellen Negativinformationen im Sinne der Theorie der Quellenglaubwürdigkeit dar. Das bedeutet, dass ein Informationsverlust eingetreten ist. Der Prozessgewinn ist somit ein Pyrrhussieg für Produzenten guter Qualität und Endverbraucher mit Qualitätsbewusstsein.

Natürlich sind das auch schlechte Nachrichten für Stiftung Warentest. Die Verkaufszahl der Test-Zeitschrift könnte sinken. Da die Glaubwürdigkeit des Testsiegels insbesondere bei jenen gelitten hat, die sich an die Negativschlagzeilen erinnern, sind insbesondere Endverbraucher mit Qualitätsbewusstsein betroffen. Es sind nämlich jene Endverbraucher, die solche Zeitschriften wie die Test-Zeitschrift abonnieren. Sie könnten ihre Abonnements deswegen kündigen. Ob solche Reaktionen existieren, ist zwar nicht bekannt, bekannt ist aber, dass sich die Verkaufszahl der Testzeitschrift seit 1991 fast halbiert hat. Aufgrund dessen muss Stiftung Warentest handeln. Eine Möglichkeit ist, die Glaubwürdigkeit des Testsiegels zu erhöhen. Eine andere Möglichkeit ist, das Testsiegel zu bewerben.

Wir finden wir heraus, dass die Glaubwürdigkeit des Testsiegels dadurch erhöht werden kann, dass Bezugsgrößen angegeben werden, die Unabhängigkeit von Stiftung Warentest gestärkt wird und Labormethoden verwendet werden. Diese Ergebnisse decken sich mit empirischen Befunden in anderen Untersuchungen. In Bezug auf ein beliebiges Qualitätssiegel stellen wir fest, dass die Unabhängigkeit des Emittenten entscheidend für die Glaubwürdigkeit eines Qualitätssigels ist. Das gilt auch für die Transparenz der Methodik und die Verwendung von Labormethoden. Unabhängigkeit und Labormethoden sind entscheidend für die Glaubwürdigkeit sowohl in Bezug auf das Testsiegel als auch in Bezug auf ein beliebiges Qualitätssiegel.

Ferner finden heraus, dass sich die Glaubwürdigkeit eines Qualitätssiegels im Allgemeinen aus zwei Hauptquellen speist: aus harten Fakten in Bezug auf den Emittenten und das Qualitätssiegel (Unabhängigkeit, Transparenz, Labormethoden und Aktualität) und aus der Gegenwärtigkeit des Qualitätssiegels (der aktiven und der passiven Gegenwärtigkeit sowie der Präsenz des Qualitätssiegels auf vielen Verpackungen). Interessant ist, dass sich die Relevanzen beider Hauptquellen kaum unterscheiden. Wohingegen die harten Fakten vollständig durch den Emittenten eines Qualitätssiegels kontrolliert werden können, kann die Gegenwärtigkeit eines Qualitätssigels nur teilweise durch den Emittenten kontrolliert werden. Immerhin kann ein Qualitätssiegel beworben werden.

Insgesamt sind das gute Nachrichten, sowohl für Emittenten als auch für Produzenten guter Qualität. Bildlich gesprochen: Sie sitzen im selben Boot. Emittenten sollten sich um ihre harten Fakten kümmern. Zudem sollten sie ihre Qualitätssiegel bewerben, wenn diese nur selten verwendet werden. Produzenten guter Qualität sollten Qualitätssiegel auf ihren Verpackungen abdrucken, um eine höhere Zahlungsbereitschaft zu generieren. Dadurch wären Quali-

tätssiegel präsenter. Davon würden auch Emittenten profitieren. Da auch Endverbraucher mit Qualitätsbewusstsein durch die zusätzlichen Informationen auch gewännen, ergäbe sich eine Win-win-win-Situation.

Executive Summary (English)

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 labelling an artificial flavouring as a natural flavouring. *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/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 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 analyse 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, 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* has to act. 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. For any quality label, we find that the independence of an awarding organization is essential for trusting in the credibility of a quality label. The same holds true for methodological transparency and laboratory methods. Independence and laboratory methods are found to be essential for the credibility for both the Test label and any quality label.

We also find for any quality label that the credibility is determined by two main sources: hard facts concerning the awarding organization and the label (independence, transparency, laboratory, and actuality) and the presence of the label (active presence, passive presence, and presence on the packaging of many products). An interesting aspect is that the importance of the presence of a label is smaller but not substantially smaller than the hard facts. Hard facts can be controlled by awarding organizations. The presence of a label cannot be completely controlled by awarding organizations. However, a label can be advertised.

This is good news for both awarding organizations and high-quality producers. Loosely speaking, they are in the same boat. Awarding organizations should care about their hard facts. In addition, they should advertise their quality labels if their labels were seldom used. High-quality producers should use these labels on the packaging of their products because there would be more willingness to pay. In consequence, these labels would be more present. From that, awarding organizations would gain. Because also quality-conscious consumers would gain from the additional information, a win-win-win situation could be created.

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 labelling an artificial flavouring as a natural flavouring. *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/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 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 analyse 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 increased by providing reference values to the tests, strengthening the independence of *Stiftung Warentest*, and using laboratory methods in the tests. For the most part, the same holds true for any quality label.

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 result are summarized and discussed. We conclude in the last section.

The appendix contains an English translation of the questions of the questionnaire (originally written in German, see also the appendix).

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/West* 1991) or the information on a source (see *Klebba/Unger* 1983). Negative information is regularly found to decrease source credibility, and positive information is usually found to increase source credibility (see *Klebba/Unger* 1983). The more credible a source is, the more persuasive it will be (see *Pornpitakpan* 2004; *Gierl/Stich/Strohmayr* 1997; *Sternthal/Phillips/Dholakia* 1978).

2.1. 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/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/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 (in descending order) the awarding organization is independent (1), reference values are given (2), the methodology is transparent (3), laboratory methods are used (4), and negative headlines are absent (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.

2.2. 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/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 consumer, 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. There are however 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 flavourings was wrong and misleading (*Ruhwinkel* 2014).

Hypothesis 2.1: (a) More (perceived) independence can increase the credibility of the *Test* label. (b) The same holds true for any quality label.

Hypothesis 2.2: (a) The (perceived) presence of referenve values is positively related to the credibility of the *Test* label. (b) The same holds true for any quality label.

Hypothesis 2.3: (a) More (perceived) transparency can increase the credibility of the *Test* label. (b) The same holds true for any quality label.

Hypothesis 2.4: (a) The (perceived) presence of laboratory methods is positively related to the credibility of the *Test* label. (b) The same holds true for any quality label.

3. Experimental design and procedures

To test our hypotheses, we use experimental methods.

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.

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

Table 1: Treatments.

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(BASELINE ∪ RECALL) = ½ and P(INFORMATION ∪ AMPLIFICATION) = 1 – P(BASELINE ∪ RECALL) = ½.

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 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 prices were collected. Three winners did not show up (retrieved March 24, 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.

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 (in 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.

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

Table 2: Treatments and participants.

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.

4.1. Do negative headlines undermine the credibility of a quality seal?

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 analysed by comparing BASELINE to RECALL or INFORMATION. By comparing BASELINE to RECALL, the natural treatments are analysed. 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 analysed | 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 analysed | 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 analysed | one-sided unpaired *t*-test: $p = 0.000$ with equal variances and $p = 0.003$ with unequal vari-

ances; 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.

4.2. 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 analysed 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.

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

Table 3: Variables, descriptions and summary statistics.

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.

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

Table 4: Estimated values for both models.

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 (a), 2.2 (a), and 2.4 (a) are supported. More (perceived) independence can increase the credibility of the *Test* label. The (perceived) presence of referenve values and laboratory methods is positively related to the credibility of the *Test* label. Hypothesis 2.3 (a) is not supported. More (perceived) transparency cannot increase the credibility of the *Test* label.

Whether the same holds true for any quality label can be examined by inspecting Figure 1. The participants were asked when they trust in a quality label. In particular, they were asked to what extent the depicted attributes influence the credibility of a quality label. For the participants, the independence of an awarding organization is essential for trusting in the credibility of a quality label. The same holds true for methodological transparency and laboratory methods. Hypotheses 2.1 (b), 2.3 (b), and 2.4 (b) are supported. Reference values are not that important. Hypothesis 2.2 (b) is not supported.

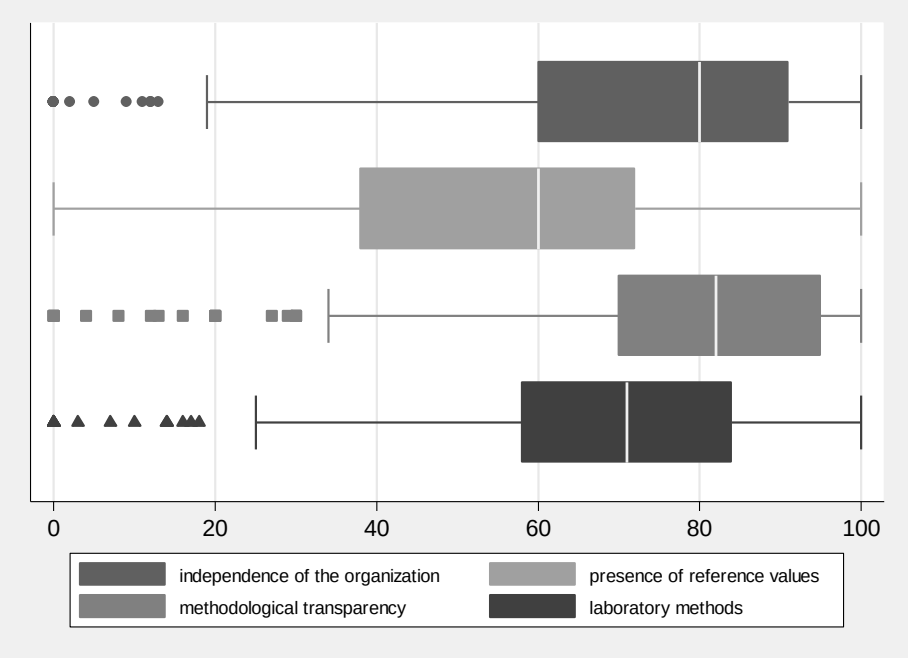

Figure 1: Attributes influencing the credibility of any quality label.

Independence and laboratory methods are found to be essential for the credibility for both the *Test* label and any quality label. That is why we focus on these attributes in the next section.

Using a principal component analysis, the credibility of a quality label can be further investigated. It is found to be determined by two main sources: hard facts concerning the awarding organization and the label (1) and the presence of the label (2). The Kaiser-Meyer-Olkin measure of sampling adequacy is 0.7196. Therefore, the sampling adequacy is acceptable. In total, 60.41 percent of the variance is explained by both main sources. The main sources, the criteria and Cronbach's alpha are given in Table 5. Because the presence of reference values and the presence of negative headlines are unique criteria, these two criteria are excluded.

source	criterion (load)	Cronbach's alpha
(1) hard facts	independence of the awarding organization (0.7858) methodological transparency (0.7811) laboratory methods (0.6907) date of the awarding (0.6552)	0.7137
(2) presence of the label	active presence on the media (0.8619) passive presence on the media (0.8279) presence on the packaging of many products (0.7460)	0.7532

Table 5: Main sources of credibility.

(1) Hard facts can be controlled by awarding organizations. They decide about their funding scheme, that is, their independence. Awarding organizations also chooses their methods, and they are free to publish them. The same holds true for the date of the awarding. (2) The presence of the labels cannot be completely controlled by awarding organizations. However, the labels can be advertised, that is, the first criterion is certainly under their control.

An interesting aspect is that the importance of the presence of a label is smaller but not substantially smaller than the hard facts. This is shown in Figure 2.

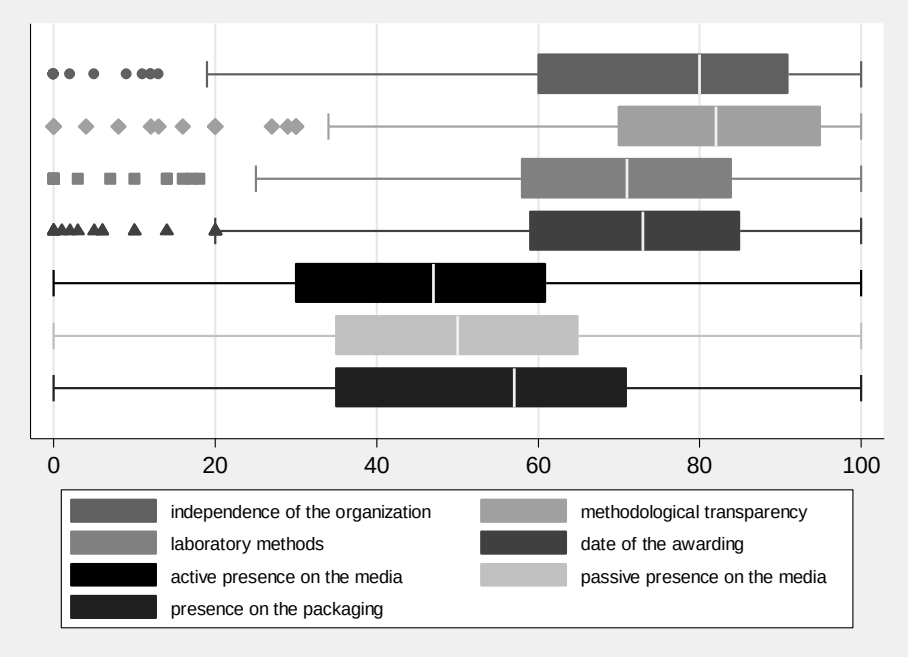

Figure 2: Importance of the criteria generating the main sources.

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 (in 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 analysed 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 ex-

cessive 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. For any quality label, we find that the independence of an awarding organization is essential for trusting in the credibility of a quality label. The same holds true for methodological transparency and laboratory methods. Independence and laboratory methods are found to be essential for the credibility for both the Test label and any quality label.

We also find for any quality label that the credibility is determined by two main sources: hard facts concerning the awarding organization and the label (independence, transparency, laboratory, and actuality) and the presence of the label (active presence, passive presence, and presence on the packaging of many products). An interesting aspect is that the importance of the presence of a label is smaller but not substantially smaller than the hard facts. Hard facts can be controlled by awarding organizations. The presence of a label cannot be completely controlled by awarding organizations. However, a label can be advertised.

This is good news for both awarding organizations and high-quality producers. Loosely speaking, they are in the same boat. Awarding organizations should care about their hard facts. In addition, they should advertise their quality labels if their labels were seldom used. High-quality producers should use these labels on the packaging of their products because there would be more willingness to pay. In consequence, these labels would be more present. From that, awarding organizations would gain. Because also quality-conscious consumers would gain from the additional information, a win-win-win situation could be created.

For further research, we suggest an analysis of the perception of the hard facts concerning awarding organizations. Do (quality-conscious) consumers (better) know the hard facts, or is there a gap between perception and reality? If there were a systematic gap, what could be

done in order to facilitate the win-win-win situation from above? Furthermore, we suggest an analysis of consumers' preferences regarding quality. Do the quality criteria of quality labels correspond to consumers' quality preferences? Only if there were no gap, a consumer could unconditionally follow the advice given by a quality label. From these analyses, further implications for the practice could be deduced.

References

Akerlof, G. A. (1970): The Market for "Lemons": Quality Uncertainty and the Market Mechanism, in: Quarterly Journal of Economics, Vol. 84, No. 3, pp. 488–500.

comScore (2014): Anzahl der Smartphone-Nutzer in Deutschland in den Jahren 2009 bis 2014, Online Data, Statista, Hamburg: http://de.statista.com/statistik/daten/studie/198959/umfrage/anzahl-der-smartphonenutzer-in-deutschland-seit-2010/ (accessed on March 24, 2015).

Darby, M R./Karni, E. (1973): Free Competition and the Optimal Amount of Fraud, in: Journal of Law and Economics, Vol. 16, No. 1, pp. 67–88.

DiNardo, J. (2008): Natural Experiments and Quasi-Natural Experiments, in: Durlauf, S. N./Blume, L. E. (Eds.): New Palgrave Dictionary of Economics, London, pp. 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: http://www.grieger-cie.de/Marktforschung/Guetesiegel-in-Deutschland-2013.pdf (accessed on March 24, 2015).

Eisend, M. (2006a): Source Credibility Dimensions in Marketing Communication: A Generalized Solution, in: Journal of Empirical Generalisations in Marketing Science, Vol. 10, No. 2, pp. 1–33.

Eisend, M. (2006b): Source Credibility in Marketing Communication: A Meta-Analysis, in: Marketing: Journal of Research and Management, Vol. 2, No. 1, pp. 43–60.

eMarketer (2014): Prognose zum Anteil der Smartphone-Nutzer weltweit von 2013 bis 2018, Online Data, Statista, Hamburg: http://de.statista.com/statistik/daten/studie/321204/umfrage/prognose-zum-anteil-der-smartphone-nutzer-weltweit/ (accessed on March 24, 2015).

Epp, A./Michalski, B./Banasiak, U. (2010): Pflanzenschutzmittel-Rückstände in Lebensmitteln: Die Wahrnehmung der deutschen Bevölkerung – Ein Ergebnisbericht, Berlin: http://www.bfr.bund.de/cm/350/pflanzenschutzmittel_rueckstaende_in_lebensmitteln.pdf (accessed on March 24, 2015).

Gierl, H./Stich, A./Strohmayr, M. (1997): Einfluß der Glaubwürdigkeit einer Informationsquelle auf die Glaubwürdigkeit der Information, in: Marketing: Zeitschrift für Forschung und Praxis, Vol. 19, No. 1, pp. 27–31.

Grossman, S. J. (1981): The Informational Role of Warranties and Private Disclosure about Product Quality, in: Journal of Law and Economics, Vol. 24, No. 3, pp. 461–483.

Haley, E. (1996): Exploring the Construct of Organization as Source: Consumers' Understandings of Organizational Sponsorship of Advocacy Advertising, in: Journal of Advertising, Vol. 25, No 2, pp. 19–35.

Hildenbrand, A./Kühl, R. (2014): Ritter Sport und Stiftung Warentest: Informationsdefizite überwinden, in: Wirtschaftsdienst, Vol. 94, No. 3, pp. 217–220.

Javalgi, R. G./Traylor, M. B./Gross, A. C./Lampman, E. (1994): Awareness of Sponsorship and Corporate Image: An Empirical Investigation, in: Journal of Advertising, Vol. 23, No. 4, pp. 47–58.

Klebba, J. M./Unger, L. S. (1983): The Impact of Negative and Positive Information on Source Credibility in a Field Setting, in: Advances in Consumer Research, Vol. 10, No. 1, pp. 11–16.

MGH (2011): QR Code Usage and Interest Survey, Research Report, MGH, Owings Mills: http://mghus.com/files/6213/5292/5743/QRcodeStats3.23.11.pdf (accessed on March 24, 2015).

Moussa, S./Touzani, M. (2008): The Perceived Credibility of Quality Labels: A Scale Validation with Refinement, in: International Journal of Consumer Studies, Vol. 32, No. 5, pp. 526–533.

Nelson, P. (1970): Information and Consumer Behavior, in: Journal of Political Economy, Vol. 78, No. 2, pp. 311–329.

Nelson, P. (1974): Advertising as Information, in: Journal of Political Economy, Vol. 82, No. 4, pp. 729–754.

Nestlé (2012): Das is(s)t Qualität: Auszüge aus der Nestlé-Studie 2012, Frankfurt: http://www.nestle.de/verantwortung/nestle-studie/2012/ (accessed on March 24, 2015).

Nielsen (2012): Bei welchen Gelegenheiten haben Sie bisher QR-Codes gescannt?, Online Data, Statista, Hamburg: http://de.statista.com/statistik/daten/studie/243059/umfrage/umfrage-zu-den-gruenden-der-bisherigen-nutzung-von-qr-codes/ (accessed on March 24, 2015).

Pornpitakpan, C. (2004): The Persuasiveness of Source Credibility: A Critical Review of Five Decades' Evidence, in: Journal of Applied Social Psychology, Vol. 34, No. 2, pp. 243–281.

Ruhwinkel, S. (2014): Streit ums Schokoladen-Aroma: Gericht verbietet Äußerungen eines Warentestinstituts, Press Release, Landgericht I, München: https://www.justiz.bayern.de/gericht/lg/m1/presse/archiv/2014/04192/ (accessed on March 24, 2015).

Schumann, D. W./Hathcote, J. M./West, S. (1991): Corporate Advertising in America: A Review of Published Studies on Use, Measurement, and Effectiveness, in: Journal of Advertising, Vol. 20, No. 3, pp. 35–56.

Sternthal, B./Phillips, L. W./Dholakia, R. (1978), The Persuasive Effect of Source Credibility: A Situational Analysis, in: Public Opinion Quarterly, Vol. 42, No. 3, pp. 285–314.

Stiftung Warentest (2011): Satzung, Berlin: https://www.test.de/unternehmen/stiftungsgremien/satzung/ (accessed on March 24, 2015).

Stiftung Warentest (2013): Licensing Agreement, Berlin: https://www.ral-logolizenz.com/fileadmin/logolizenz/pdf/RAL_Logolizenz_Conditions_of_Contract_Attachement_141205.pdf (accessed on March 24, 2015).

Stiftung Warentest (2014a): Objective and Independent, Berlin: https://www.test.de/filestore/infoflyer-en.pdf?path=/14/61/f0fa4db7-b54d-443a-943f-aec3c422eb2d-file.pdf&key=65B703CB6A720C70F1BD8264B964B22469EB423F (accessed on March 24, 2015).

Stiftung Warentest (2014b): Testablauf, Berlin: https://www.test.de/unternehmen/testablauf/ (accessed on March 24, 2015).

Stiftung Warentest (2014c): Verkaufsentwicklung, Berlin: https://www.test.de/filestore/auflage-test-gross.jpg?path=/11/a7/18ecaab9-1a02-4997-be01-a55372005ca1-file.jpg&key=88D0202AB0DF7A82C9BE92F2ABBC5C316582A011 (accessed on March 24, 2015).

Stiftung Warentest (2014d): test.de, Berlin: https://www.test.de/filestore/testde-gross.jpg?path=/91/a0/902f2895-b38b-44ea-a728-32b00ff512a1-file.jpg&key=FEC02405A07AB3FE0D417B189EB100501EA4B1C7 (accessed on March 24, 2015).

Tomorrow Focus AG (2014): Wie oft nutzen Sie mobiles Internet mit Ihrem Mobiltelefon?, Online Data, Statista, Hamburg: http://de.statista.com/statistik/daten/studie/167623/umfrage/nutzungshaeufigkeit-des-mobilen-internets-ueber-das-mobiltelefon/ (accessed on March 24, 2015).

Verbraucherzentrale Bundesverband (2008): Verbraucherschutz in Deutschland: Was meinen die Verbraucher?, Berlin: http://www.vzbv.de/sites/default/files/mediapics/verbraucherbefragung_2008.pdf (accessed on March 24, 2015).

Appendix A

1. How important are quality labels to you …
   a. in general?
   b. when you buy foods?
2. When do you trust in a quality label? In particular, to what extent do the following attributes influence the credibility of a quality label?
   a. Independence of the awarding organization: The awarding organization is economically independent.
   b. Presence of reference values: A product is compared to other products.
   c. Methodological transparency: The conditions, under which the quality label is obtained, are known.
   d. Laboratory methods: The results comply with scientific standards.
   e. Absence of negative headlines: There is no negative information on the quality label.
   f. Date of the awarding: The quality label is up to date.
   g. Presence on the packaging of many products: The quality label is often used.

h. Active presence on the media: The quality label is advertised.
i. Passive presence on the media: The quality label is used for advertising.

3. Do you know the quality label of *Stiftung Warentest* (*Test* label)?
4. Do you remember any headlines or reporting on the *Test* label or *Stiftung Warentest* from the last six months?
5. What kind of reporting do you remember?
   a. I mainly remember negative reporting on *Stiftung Warentest*.
   b. I mainly remember neutral headlines on *Stiftung Warentest*.
   c. I mainly remember positive headlines on *Stiftung Warentest*.
6. Please read the following headlines.
   a. "*Ritter Sport* beats *Stiftung Warentest* in court " (Die Welt 2014-01-13)
   b. "*Ritter Sport* inflicted a severe defeat on *Stiftung Warentest*" (Frankfurter Allgemeine Zeitung 2014-01-13)
   c. "Controversial tests put pressure on *Stiftung Warentest*" (Die Welt 2014-05-04)
7. Do you trust in *Stiftung Warentest*? In particular, how credible is the *Test* label for you?
8. To what extent do you agree with the following statements on the *Test* label?
   a. Laboratory methods are used to obtain a test result. *Stiftung Warentest* works scientifically.
   b. *Stiftung Warentest* is independent. They can freely (without economic pressure) make decisions.
   c. With the help of the *Test* label, I can better compare a product to other products. There are enough reference values.
   d. The methodology behind the *Test* label is transparent. I know the conditions under which a test result is obtained.
9. What is your sex?
10. How old are you?
11. Are you a student at Justus Liebig University Giessen?
12. Which department do you belong to?

Appendix B

1. Wie wichtig sind Ihnen Qualitätssiegel beim Kauf …
   a. im Allgemeinen?
   b. von Lebensmitteln?

2. Wann vertrauen Sie auf ein Qualitätssiegel? Das heißt, inwieweit beeinflussen die folgenden Attribute die Glaubwürdigkeit eines Qualitätssiegels?
    a. Unabhängigkeit des Ausstellers: Der Aussteller ist wirtschaftlich unabhängig.
    b. Angabe von Bezugsgrößen: Ein Produkt wird mit anderen Produkten verglichen.
    c. Transparenz der Methodik: Die Rahmenbedingungen, unter denen das Qualitätssiegel vergeben wird, sind bekannt.
    d. Verwendung von Labormethoden: Die Ergebnisse genügen wissenschaftlichen Ansprüchen.
    e. Abwesenheit von Negativschlagzeilen: Es liegen keine negativen Informationen über das Qualitätssiegel vor.
    f. Datum der Vergabe: Das Qualitätssiegel ist aktuell.
    g. Präsenz auf der Verpackung vieler Produkte: Das Qualitätssiegel wird oft verwendet.
    h. Aktive Präsenz in den Medien: Das Qualitätssiegel wird beworben.
    i. Passive Präsenz in den Medien: Mit dem Qualitätssiegel wird geworben.
3. Ist Ihnen das Qualitätssiegel der Stiftung Warentest (Testsiegel) bekannt?
4. Erinnern Sie sich an Schlagzeilen oder Berichterstattungen über das Testsiegel oder Stiftung Warentest aus den letzten sechs Monaten?
5. An welche Art der Berichterstattung erinnern Sie sich?
    a. Ich erinnere mich überwiegend an negative Berichterstattungen über Stiftung Warentest.
    b. Ich erinnere mich überwiegend an neutrale Berichterstattungen über Stiftung Warentest.
    c. Ich erinnere mich überwiegend an positive Berichterstattungen über Stiftung Warentest.
6. Bitte lesen Sie die folgenden Schlagzeilen.
    a. „Ritter Sport siegt gegen Stiftung Warentest“ (Die Welt 2014-01-13)
    b. „Schwere Niederlage für Warentest gegen Ritter Sport“ (Frankfurter Allgemeine Zeitung 2014-01-13)
    c. „Umstrittene Tests bringen Warentest unter Druck“ (Die Welt 2014-05-04)
7. Vertrauen Sie auf Stiftung Warentest? Das heißt, wie glaubwürdig ist das Testsiegel für Sie?
8. Inwieweit stimmen Sie den folgenden Aussagen über das Testsiegel zu?
    a. Es werden Labormethoden verwendet, um ein Testurteil zu erzeugen. Stiftung Warentest arbeitet wissenschaftlich.

b. Stiftung Warentest ist unabhängig. Sie kann Entscheidungen frei (ohne wirtschaftlichen Zwang) treffen.
c. Mithilfe des Testsiegels kann ich ein Produkt besser mit anderen Produkten vergleichen. Es gibt genügend Bezugsgrößen.
d. Die Methodik hinter dem Testsiegel ist transparent. Ich weiß, unter welchen Rahmenbedingungen ein Testurteil zustande kommt.

9. Welches Geschlecht haben Sie?
10. Wie alt sind Sie?
11. Studieren Sie an der Justus-Liebig-Universität Gießen?
12. Welchem Fachbereich gehören Sie an?