How Trustworthiness Seals Can Highlight Information and Influence Decisions

Rick Hayes-Roth
Professor, Information Sciences Department
Naval Postgraduate School
Monterey, CA 93943
July 21, 2011

Abstract
People making decisions require information about their options. Across a wide range of tasks, they receive information from diverse sources, such as the Web, print, radio and television. As the Internet Age progresses, each decision maker must increasingly assess the credibility of the author, the credibility of the evidence the author cites, and the credibility of the publisher in order to gauge the credibility of the information. We wanted to determine whether publishers and authors could affix “trustworthiness seals” to stated claims to increase their persuasiveness. We created pairs of descriptions for comparable options that employed no seals, some weak generic seals, and some strong seals guaranteeing veracity. Experimental subjects rated their preferences for each option promoted with an advertisement bearing some of these seals and also made forced choices between pairs of comparable options. The results show that all seals have a significant effect on perceived attractiveness of options and that the strong seals produce the greatest increase. This study suggests that authors, advertisers and publishers can significantly boost their effectiveness through an independent validation and by guaranteeing the truthfulness of their claims. This potential can pave the way for market mechanisms that reward truth-telling and improved tools for filtering information based on information credibility.

Introduction

Even though lying has been going on through all of recorded history, some worry the Internet Age has unleashed widespread “information pollution” (Hayes-Roth, 2011). The reasons and motives for publishing bogus information are many. First, companies have agendas that can benefit from propaganda and well-chosen misrepresentations (Jensen, 2011). The same is true of politicians in political campaigns. Many websites make money by getting eyeballs and directing them to paying advertisers. Success in these endeavors means getting eyeballs or followers or adherents. There is little or no penalty for lying. In a world where results are measured only by dollars or votes, honesty seems quaint and mostly irrelevant (Denning, 2011).

A number of people and organizations have dedicated their efforts to ferreting out truth, publicizing falsehoods, and promoting the public interest in making decisions based on facts rather than fictions. Some of these organizations include PolitiFact.com, FactCheck.org, Snopes.com, and TruthSeal.org. Traditionally, many journalists and news organizations focused on these objectives as well. However, traditional journalism has
How Seals Increase Ratings and Preferences

waned dramatically as the advertising dollars found better venues for selling their products, services, and candidates. The ad dollars mostly pay for information production and dissemination, and they had little need to stick with truth tellers as the truth tellers lost audience to entertainment and infotainment. As an example, GE which owned NBC until it sold it to Comcast, saw NBC News as a rounding error in its entertainment business, which itself was a tiny proportion of the whole business portfolio. Some serious scholars worry that an unbridled focus on the almighty buck rather than more traditional values such as honesty could jeopardize the foundations of both democracy and capitalism (Fukuyama, 1995; Zak, 2008).

The Web, on the other hand, has brought forth a great deal of well grounded information, including well polished Wikipedia articles, YouTube tutorials, and on-line tools for statistics and finance. In these cases, one can quickly find substantiated information by following links to it from other credible sites or using search engines that use such methods to rank pages by estimated information quality. In fact, the linking between one statement and another often directly expresses some grounding for the statement, such as a source article or relevant evidence. When we try to assess the credibility of Web information, we often informally look at the local network of links to determine whether the information of interest rests on credible sources and whether it is trusted by others whom we consider trustworthy. In this way, we attempt to judge the trustworthiness of the information.

We see a potential opportunity to generalize these ideas to make information routinely assessable in this way. Specifically, we envision an extensively linked web of information where a new distinguished type of link is employed explicitly to address the question of how credible a statement is. Specifically, we anticipate that each claim can be linked immediately to a particular sort of meta-data that describes the evidence that grounds the claim, the current truth value status of the claim, and the rationale by which that status was determined. If such meta-data were widely available, a reader of information could easily access this truth value meta-data for any claim, perhaps by mousing over an icon affixed to that claim signifying the truth value status.

For such truth value meta-data to become ubiquitous, authors and publishers would need to see an incentive, and consumers would need to trust the procedures employed. Thus, independent third-parties would be necessary to implement the procedures that could provide authors and publishers an incentive while maintaining the trust of consumers. Regardless of what business provided these services, the need for incentives would require that some icon signifying trustworthiness could markedly increase consumers’ trust and acceptance of the marked statements.

There are a number of trademarked icons and brands that have played some roles similar to marking the truth value of claims, but none has done that directly. Familiar brands that provide some degree of “trustworthiness seals” include Good Housekeeping, Better Business Bureau, Verisign, and Underwriters Laboratory. Each seal indicates something specific, but the average consumer doesn’t know precisely what it signifies. The licensors of those seals and trademarks provide a specific service to the companies who affix the seals to their products or services. In any case, while these seals and brands are familiar, they do not intend to signify sincerity, honesty or truth of the claims made.
by the sellers of products or services. They usually represent just some type of assurance or guarantee, often quite nebulous. Furthermore, the selling of political candidates defines a third category that has no readily recognizable seals in use.

While the potency and familiarity of seals can increase as the result of brand-building campaigns, we want to assess whether trustworthiness seals can have a marked effect on sales efforts even if the seals are brand new. Market research in Denmark has shown a major increase on click-through sales of products from web sites carrying an “E-mark” seal signifying ethical e-business practices (Lund, 2010). We can find no other research on the impact of seals on customer, consumer, or citizen preferences and decision-making. Some research has been published on the history of various efforts to license seals, but it doesn’t include evidence on the effectiveness of seals to increase customer preferences and choices.

We wanted to investigate the impact of various seals on consumer preferences and decisions across three categories of options: products (P), services (S), and political candidates (C). We used three experimental conditions regarding seals. In the first case, no seals (NS) were used. In the second case, generic seals (GS) were used. Generic seals were licensed so that advertisers who wanted to express “sincerity” or “good intentions” could affix a seal. Nothing more was required for GS use, and nothing more was implied. In the final case, we described a TruthSeal™ (TS) mark of truth. One could license a TS seal only by providing the licensor credible evidence of the truth of the claim and by guaranteeing the truth of the claim. If that guaranteed truth was overturned, the challenger who presented falsifying data would receive a bounty and the guarantor would pay an additional penalty. Thus, the appearance of a TS implied a degree of increased credibility and commitment.

We wanted to test how potential customers or voters would assess two comparable options. In each pair of similar options, corresponding claims could be bear GS or TS marks or have no trustworthiness marking (NS). Presumably, TS-marked items would be judged more credible, while GS items would have a smaller advantage, if any, over NS items.

Our experiment was designed to allow us to assess the impact of each of the kinds of marks, to compare the effect each type of mark had on preferences and decisions, and to determine the relative effectiveness of each type of mark.

Hypotheses

We are most interested in assessing how various types of trustworthiness seals affect the perceived attractiveness of decision options. If people are assessing attractiveness based on descriptive information about an option, they ought to weight credible information more highly. A trustworthiness seal highlights the information it adorns and calls attention to the questions of evidence, truthfulness, and credibility. Depending on what the reader believes about the seal, the advertiser, and the licensor of the seal, the reader might add weight to the information, treat it no differently than other information, or might even discount the information as being tainted or untrustworthy. We thought it likely that a seal that signified a guaranteed truthfulness, our TS condition,
would lead people to weight the information more heavily. We thought it likely that a seal signifying little more than “goodness”, our GS condition, would have a mild positive effect. Both of these effects would be apparent in subjects preferring the alternative in any pair marked by a TS or GS in comparison to the option bearing NS information. Thus, our basic hypothesis is that Choices with TS > Choices with GS > Choices with NS. This shorthand is read as “The options marked with TruthSeals would be preferred to the options marked with generic seals, and the GS-marked options would be preferred to the unmarked options. These hypotheses are labeled as H1 through H3:

H1: TS options are preferred to GS options.
H2: GS options are preferred to NS options.
H3: TS options are preferred to options bearing either NS or GS.

In addition, we thought decision makers would respond similarly regardless of the category under consideration. Specifically, we expected the same preference behavior to occur in each of our three categories, products (P), services (S), and political candidates (C). So if we interpret the hypotheses H1-H3 as being across all categories, we can specialize each of these hypotheses to each of the three categories, P, S and C. Thus, we have an additional nine hypotheses, enumerated below:

P-H1: TS options are preferred to GS options when assessing products.
P-H2: GS options are preferred to NS options when assessing products.
P-H3: TS options are preferred to options bearing either NS or GS when assessing products.
S-H1: TS options are preferred to GS options when assessing services.
S-H2: GS options are preferred to NS options when assessing services.
S-H3: TS options are preferred to options bearing either NS or GS when assessing services.
C-H1: TS options are preferred to GS options when assessing political candidates.
C-H2: GS options are preferred to NS options when assessing political candidates.
C-H3: TS options are preferred to options bearing either NS or GS when assessing political candidates.

**Methods**

**Participants**

We used two pools of subjects. 106 subjects were obtained from a commercial company (United Sample) that provides panels for survey research. They were commissioned to provide at least 50 US voters in each of two groups, with ages and genders matching the US voter population, chosen at random across the US. Our second
group was obtained via email solicitations through colleagues to students, friends, and associates who asked for volunteers. Subjects in the first pool were paid for participation. Subjects in the second pool were offered a random chance to win a token sweepstakes prize of $100. All surveys were administered over the Internet using SurveyMonkey. The results were nearly identical across the two pools, so after preliminary analysis showed no differences, we combined the two samples into one for the purpose of analysis.

The resulting total pool included 173 respondents. Their ages and genders are as indicated below:

Table 1. Distribution of subjects by age and gender.

<table>
<thead>
<tr>
<th>Age</th>
<th>Gender</th>
<th>Male</th>
<th>Female</th>
<th>Grand Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>18-29</td>
<td></td>
<td>34</td>
<td>22</td>
<td>56</td>
</tr>
<tr>
<td>30-49</td>
<td></td>
<td>29</td>
<td>25</td>
<td>54</td>
</tr>
<tr>
<td>Over 50</td>
<td></td>
<td>39</td>
<td>24</td>
<td>63</td>
</tr>
<tr>
<td>Grand Total</td>
<td></td>
<td>102</td>
<td>71</td>
<td>173</td>
</tr>
</tbody>
</table>

**Design and procedure**

Each subject was asked to complete a survey that asked him or her to consider information about decision options and then answer questions. Each option was presented and then the subject rated the attractiveness on a 7-point scale with 1 being “Least Appealing” and 7 being “Most Appealing.”. After each option within a pair was presented and rated, the subject was shown both options side by side and asked to make a forced choice between the two. Three pairs of products were presented and assessed first. Then three pairs of services were rated, and finally three pairs of candidates were judged.

Two different surveys were created, which we’ll call A and B. Half of the paid subjects were given the A instrument, and half the B instrument. The instruments were identical except that every GS in A was replaced by TS in B, and every TS in A was replaced by GS in B. This controlled for the possibility that the particular items marked with a TS might be inherently more appealing than those marked with a GS, or vice versa. There were no differences between the A and B groups with respect to the hypotheses. The number of subjects in the second group wasn’t under control of the experimenter, who merely asked colleagues to distribute either the A or B survey to their students, colleagues and associates. As a result, the second pool provided 52 and 15 additional subjects for the A and B groups, respectively. In the end, we received and analyzed data from all the survey respondents, numbering 104 and 69 in the two respective groups. No differences were apparent arising from the assignment to the A or B instrument.

A pilot study was done to determine how to keep the survey concise, so that subjects would give attention to all information presented. In the pilot study, we described all of the various types of seals at the beginning and then presented just side-by-side pairs of advertisements. This overloaded subjects, some of whom admitted they skimmed the information and found the assessments complex.
In response to that feedback, we made the survey more concise and serial to reduce the cognitive load. In the resulting instrument, a little information was presented at the beginning and one option was presented at a time. Information about a novel type of seal was introduced just when the seal first appeared. After each option was rated on the 7-point scale of appealingness, the pair of options was presented together, side-by-side. We then asked the subject to make a forced choice between the two. This appeared to make the task subjectively less taxing, and no subjects reported difficulties.

We used two different types of generic seals. One was called the Gold Star Shield for use with products or services. Another type of generic seal was called the GoodCitizens Seal, which was applicable only to political campaigns and their candidates. TruthSeals were applicable to all categories. Because this meant that generic seals would be introduced twice collectively, a brief reminder about the TruthSeal was provided during the candidate evaluations to refresh the subject who was about to compare the newly introduced GoodCitizens Seal and the previously introduced TruthSeal.

The full survey instrument for the A group is attached as Appendix 1. The two instruments are practically identical, except the items marked by GS and TS in A are marked as TS and GS in B, respectively. Also, the texts used to introduce the GS and TS seals are moved to the appropriate place when the seals are first used. There are no other differences between the A and B instruments.

We cited a number of other information sources in the instruments, such as newspaper ratings, Facebook “Like” counts, and consumer ratings. These were practically identical in each option of a pair and provided no apparent basis for preferring one option over the other.

The actual descriptions that carried seals in the two groups were designed to be nearly equivalent or at least comparable, but in some cases a factual statement such as product 1 is better than product 2 might be used in the advertisement for product 1. In such a case, the product 2 ad would claim general advantages or popularity to avoid contradicting something claimed to be true. The actual content might be expected to influence relative ratings, and for that reason the roles of TS and GS were switched between A and B, which would control for the impact of specific content as opposed to the effect of the seals per se.

Results

Data were used only from people who completed the surveys. There were 173 completed surveys out of 177 started. Subjects could stop taking the survey if they wished, and 2% did. Data were analyzed exclusively on a within-subject basis. When an hypothesis predicted that one item or type of item would be preferred over another, this could be assessed in two ways. We could test the null hypothesis of a 50-50 random choice between the two alternatives using a binomial test. As an alternative, we could test the null hypothesis that the difference between two ratings of appealingness would be zero using a z-test. The reliability of the preference ratings was assessed for all ratings of items of each seal type NS, GS, and TS. The Cronbach α for these three sets of items over
all subjects were 0.888, 0.886, and 0.878 respectively. All of the different analysis methods produced compatible results for all of the hypotheses. The results are summarized as follows:

All of the 12 hypotheses were confirmed at significant levels.

Decision options whose supporting claims bore any type of seal were preferred over those having no seals.

Generic seals increase the preference for the option carrying that seal over a comparable option with no seal, even when the generic seal promises essentially nothing.

TruthSeals, marking a claim as being guaranteed true, significantly increased preferences over the comparable alternatives carrying a generic seal.

TruthSeals, indicating vetted and guaranteed claims about a potential choice, significantly improve the odds that a person will select that choice.

In the following table, statistical significance is represented at four possible levels, p < .05, p < .025, p < .01, and p < .001, indicated as *, **, and *** respectively. Tests of forced choices were done using binomial tests, and tests of preference rating advantages were done using z-tests. All hypotheses were assessed using one-tailed tests. The results pertinent to the 12 hypotheses are summarized in the following table.

<table>
<thead>
<tr>
<th>Hypothesis</th>
<th>Shorthand</th>
<th>Measure</th>
<th>Variable</th>
<th>Statistic</th>
<th>Estimate</th>
<th>p (1-tailed)</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>H1, forced</td>
<td>TS:GS</td>
<td>Prefer TS&lt;sub&gt;i&lt;/sub&gt; over GS&lt;sub&gt;j&lt;/sub&gt;</td>
<td></td>
<td>B(N,p,k), N=519, p=0.5, k=333</td>
<td>TS preferred over GS</td>
<td>&lt; .001 ***</td>
<td></td>
</tr>
<tr>
<td>choice odds</td>
<td></td>
<td>for all pairs i</td>
<td></td>
<td></td>
<td>odds = 1.79</td>
<td></td>
<td></td>
</tr>
<tr>
<td>H1, ratings</td>
<td>TSr-GSr</td>
<td>Average TS item ratings - Average GS item</td>
<td>Average (TSrating&lt;sub&gt;ijk&lt;/sub&gt;) - Average</td>
<td>z(rating differences,</td>
<td>Mean difference of average ratings = 0.39</td>
<td>&lt; .001 ***</td>
<td></td>
</tr>
<tr>
<td>differences</td>
<td></td>
<td>ratings for each subject j &gt; 0</td>
<td>(GSrating&lt;sub&gt;ijk&lt;/sub&gt;) for all items i for each subject k &gt; 0</td>
<td>assumed mean =0), N=173</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>H2, forced</td>
<td>GS:NS</td>
<td>Prefer GS&lt;sub&gt;i&lt;/sub&gt; over NS&lt;sub&gt;j&lt;/sub&gt;</td>
<td></td>
<td>B(N,p,k), N=519, p=0.5, k=366</td>
<td>GS preferred over NS</td>
<td>&lt; .001 ***</td>
<td></td>
</tr>
<tr>
<td>choice odds</td>
<td></td>
<td>for all pairs i</td>
<td></td>
<td></td>
<td>odds = 2.39</td>
<td></td>
<td></td>
</tr>
<tr>
<td>H2, ratings</td>
<td>GSr-NSr</td>
<td>Average GS item ratings - Average</td>
<td>Average (GSrating&lt;sub&gt;ijk&lt;/sub&gt;) - Average</td>
<td>z(rating differences,</td>
<td>Mean difference of average ratings =</td>
<td>&lt; .001 ***</td>
<td></td>
</tr>
<tr>
<td>differences</td>
<td></td>
<td>ratings for each subject j &gt; 0</td>
<td>(NSrating&lt;sub&gt;ijk&lt;/sub&gt;) for all items i for each subject k &gt; 0</td>
<td>assumed mean =0), N=173</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
How Seals Increase Ratings and Preferences

NS item ratings for each subject j > 0 for all items i for each subject k > 0 0.65

H3, forced choice odds

| TS:GS|NS | Prefer TS_i over GS_i or NS_i for all pairs i |
|----|----|--------------------------------|----------------------------------|
| [TS_i; GS_i]NS_i => TS_i | / | [TS_i;GS_i] |
| B(N,p,k), N=1038, p=0.5, k=730 | TS preferred over GS|NS odds = 2.37 |
| < .001 | *** |

H3, ratings differences

| TSr-GSr|NSr | Average TS item ratings - Average GS & NS item ratings for each subject j > 0 |
|----|----|---------------------------------|---------------------------------|
| z(rating differences, assumed mean =0), N=173 | Mean difference of average ratings = 0.65 |
| < .001 | *** |

P-H1, forced choice odds

<table>
<thead>
<tr>
<th>P:[TS:GS]</th>
<th>Prefer TS_i over GS_i for all pairs i</th>
</tr>
</thead>
<tbody>
<tr>
<td>[TS_i;GS_i] =&gt; TS_i</td>
<td>/</td>
</tr>
<tr>
<td>B(N,p,k), N=173, p=0.5, k=118</td>
<td>TS preferred over GS odds = 2.15</td>
</tr>
<tr>
<td>&lt; .001</td>
<td>***</td>
</tr>
</tbody>
</table>

P-H1, ratings differences

<table>
<thead>
<tr>
<th>P:[TSr-GSr]</th>
<th>Average TS item ratings - Average GS item ratings for each subject j &gt; 0</th>
</tr>
</thead>
<tbody>
<tr>
<td>z(rating differences, assumed mean =0), N=173</td>
<td>Mean difference of average ratings = 0.53</td>
</tr>
<tr>
<td>&lt; .001</td>
<td>***</td>
</tr>
</tbody>
</table>

P-H2, forced choice odds

<table>
<thead>
<tr>
<th>P:[GS:NS]</th>
<th>Prefer GS_i over NS_i for all pairs i</th>
</tr>
</thead>
<tbody>
<tr>
<td>[GS_i;NS_i] =&gt; GS_i</td>
<td>/</td>
</tr>
<tr>
<td>B(N,p,k), N=173, p=0.5, k=111</td>
<td>GS preferred over NS odds = 1.79</td>
</tr>
<tr>
<td>&lt; .001</td>
<td>***</td>
</tr>
</tbody>
</table>

P-H2, ratings differences

<table>
<thead>
<tr>
<th>P:[GSr-NSr]</th>
<th>Average GS item ratings - Average NS item ratings for each subject j &gt; 0</th>
</tr>
</thead>
<tbody>
<tr>
<td>z(rating differences, assumed mean =0), N=173</td>
<td>Mean difference of average ratings = 0.54</td>
</tr>
<tr>
<td>&lt; .001</td>
<td>***</td>
</tr>
</tbody>
</table>

P-H3, forced choice odds

<table>
<thead>
<tr>
<th>P:[TS:GS</th>
<th>NS]</th>
<th>Prefer TS_i over GS_i or NS_i for all pairs i</th>
</tr>
</thead>
<tbody>
<tr>
<td>[TS_i; GS_i]</td>
<td>NS_i =&gt; TS_i</td>
<td>/</td>
</tr>
<tr>
<td>B(N,p,k), N=346, p=0.5, k=252</td>
<td>TS preferred over GS</td>
<td>NS odds = 2.68</td>
</tr>
<tr>
<td>&lt; .001</td>
<td>***</td>
<td></td>
</tr>
</tbody>
</table>

P-H3, ratings differences

<table>
<thead>
<tr>
<th>P:[TSr-GSr</th>
<th>NSr]</th>
<th>Average TS item ratings - Average GS item ratings for each subject j &gt; 0</th>
</tr>
</thead>
<tbody>
<tr>
<td>z(rating differences, assumed mean =0), N=173</td>
<td>Mean difference of average ratings =</td>
<td></td>
</tr>
<tr>
<td>&lt; .001</td>
<td>***</td>
<td></td>
</tr>
</tbody>
</table>
How Seals Increase Ratings and Preferences

<table>
<thead>
<tr>
<th>S-H1, forced choice odds</th>
<th>S: [TS:GS]</th>
<th>Prefer TS (_i) over GS (_i) for all pairs (i)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>B(N, p, k), N=173, p=0.5, k=114</td>
<td>TS preferred over GS odds = 1.93</td>
</tr>
<tr>
<td></td>
<td>z(rating differences, assumed mean =0), N=173</td>
<td>Mean difference of average ratings = 0.36</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>S-H1, ratings differences</th>
<th>S: [TSr-GSr]</th>
<th>Average TS item ratings - Average GS item ratings for each subject (j &gt; 0)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>B(N, p, k), N=173, p=0.5, k=120</td>
<td>GS preferred over NS odds = 2.26</td>
</tr>
<tr>
<td></td>
<td>z(rating differences, assumed mean =0), N=173</td>
<td>Mean difference of average ratings = 0.37</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>S-H2, forced choice odds</th>
<th>S: [GS:NS]</th>
<th>Prefer GS (_i) over NS (_i) for all pairs (i)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>B(N, p, k), N=173, p=0.5, k=120</td>
<td>GS preferred over NS odds = 2.26</td>
</tr>
<tr>
<td></td>
<td>z(rating differences, assumed mean =0), N=173</td>
<td>Mean difference of average ratings = 0.37</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>S-H2, ratings differences</th>
<th>S: [GSr-NSr]</th>
<th>Average GS item ratings - Average NS item ratings for each subject (j &gt; 0)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>B(N, p, k), N=346, p=0.5, k=237</td>
<td>TS preferred over GS</td>
</tr>
<tr>
<td></td>
<td>z(rating differences, assumed mean =0), N=173</td>
<td>Mean difference of average ratings = 0.73</td>
</tr>
</tbody>
</table>

| S-H3, forced choice odds | S: [TS:GS|NS] | Prefer TS \(_i\) over GS \(_i\) or NS \(_i\) for all pairs \(i\) |
|--------------------------|--------------|-----------------------------------------------------------------|
|                         | B(N, p, k), N=346, p=0.5, k=237 | TS preferred over GS|NS odds = 2.17 |
|                         | z(rating differences, assumed mean =0), N=173 | Mean difference of average ratings = 0.73 |

| S-H3, ratings differences | S: [TSr-GSr|NSr] | Average TS item ratings - Average GS & NS item ratings for each subject \(j > 0\) |
|---------------------------|--------------|-----------------------------------------------------------------|
|                          | B(N, p, k), N=173, p=0.5, k=101 | TS preferred over GS odds = 1.93 |
|                          | z(rating differences, assumed mean =0), N=173 | Mean difference of average ratings = 0.73 |

<table>
<thead>
<tr>
<th>C-H1, forced choice odds</th>
<th>C: [TS:GS]</th>
<th>Prefer TS (_i) over GS (_i) for all pairs (i)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>B(N, p, k), N=173, p=0.5, k=101</td>
<td>TS preferred over GS odds = 1.93</td>
</tr>
<tr>
<td></td>
<td>z(rating differences, assumed mean =0), N=173</td>
<td>Mean difference of average ratings = 0.73</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>C-H1, ratings differences</th>
<th>C: [TSr-GSr]</th>
<th>Average TS item ratings - Average GS &amp; NS item ratings for each subject (j &gt; 0)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>B(N, p, k), N=173, p=0.5, k=101</td>
<td>TS preferred over GS odds = 1.93</td>
</tr>
<tr>
<td></td>
<td>z(rating differences, assumed mean =0), N=173</td>
<td>Mean difference of average ratings = 0.73</td>
</tr>
</tbody>
</table>
How Seals Increase Ratings and Preferences

Average GS item ratings for each subject j > 0 (GSrating_{i,k}) for all items i for each subject k > 0 =0, N=173 ratings = 0.29

C-H2, forced choice odds C:[GS:NS] Prefer GS_{i} over NS_{i} for all pairs i [GS_{i}:NS_{i}] => GS_{i}/|GS_{i}:NS_{i}| B(N,p,k), N=173, p=0.5, k=119 GS preferred over NS odds = 2.20 p < .001 ***

C-H2, ratings differences C:[GSr-NSr] Average GS item ratings - Average NS item ratings for each subject j > 0 Average (GSrating_{i,k}) - Average (NSrating_{i,k}) for all items i for each subject k > 0 z(rating differences, assumed mean =0), N=173 Mean difference of average ratings = 0.65 p < .001 ***

C-H3, forced choice odds C:[TS:GS|NS] Prefer TS_{i} over GS_{i} or NS_{i} for all pairs i [TS_{i}:GS_{i}|NS_{i}] => TS_{i}/|TS_{i}:GS_{i}| B(N,p,k), N=346, p=0.5, k=241 TS preferred over GS|NS odds = 2.30 p < .001 ***

C-H3, ratings differences C:[TSr-GS|NSr] Average TS item ratings - Average GS & NS item ratings for each subject j > 0 Average (TSrating_{i,k}) - Average (GSrating_{i,k}, NSrating_{i,k}) for all items i for each subject k > 0 z(rating differences, assumed mean =0), N=173 Mean difference of average ratings = 0.61 p < .001 ***

The rating comparisons assessed within each subject the average of all the ratings for one type of item against the average ratings for the other type of item. The difference between the first and the second average would represent how many points, on the 7-point scale of least appealing to most appealing, that each subject would prefer the first type of item. For example, in comparing the ratings for TruthSeal-marked items, all of those individual items would have their ratings averaged. If the average of the Truth-Sealed ratings minus the average ratings for another type of item were positive, that subject preferred items with TruthSeals over the alternative. In this way, a within-subject preference could be computed, and the average of these differences across all subjects would be distributed as a z-statistic with a zero mean under the null hypothesis of no systematic preference. The actual means were significantly positive in each of the 3 categories and across all categories combined. Thus all of the hypotheses were confirmed on the basis of ratings.

The forced-choice comparisons were based on within subject preferences for items of the first type vs. items of the comparison type. There was exactly one such pair per category, so the N for the P, S and C categories was 173 in either comparing TS to
GS or GS to NS or TS to NS. Across all categories, we had a maximum N of 519 of forced paired comparisons. Within each category, all hypothesized preferences of TS over GS, GS over NS, and TS over either GS or NS were confirmed. The same was true when items from all categories were combined, under the assumption that comparisons of different pairs chosen from three different categories are statistically independent.

In short, all evidence supports the hypotheses that seals make a difference in how people appraise options, leading them to prefer options supported by statements marked with seals. The generic seals connoted some goodness of products, businesses, or campaigns, with vague statements as to the actual basis or implication. The TruthSeal marks were described as guaranteeing the truth of the associated claim against a penalty if disproved. That conveyed a stronger sense of trustworthiness, and this resulted in the subjects preferring the options whose descriptions included positive claims marked with TruthSeals.

**Conclusion**

We live in a world of information glut (Denning, 2006). On the one hand, we have ready access to more information than we can possibly digest or even browse. In addition, we are bombarded with advertising and promotional messages producing a veritable sense of attack or overload. People would clearly benefit if they could easily filter information based on its quality, such as automatically eliminating falsehoods or ungrounded claims or highlighting vetted statements guaranteed to be true. Because of the need and potential value, we anticipate this kind of filtering eventually will become common. That prospect depends on several factors, and this experiment studied one of the most basic questions: Will seals purporting to signify credibility actually influence people, and will people discern differences between vague and strong commitments to trustworthiness? The answers to these questions in our experiment were yes.

Whether you are promoting a product, service, or political candidate, you can greatly increase the odds your offering is rated favorably and selected over competing alternatives by providing positive statements that carry trustworthiness seals. Strong seals reflecting a vetting process and a truthfulness guarantee have significant advantages. In a two-alternative forced-choice, the use of TruthSeals generally increased odds of the associated option from 50-50 to more than 4:1, in going from an ad with no seals to an ad using TruthSeals. Even though generic seals had a positive impact also, TruthSeal-marked options had odds of 2.37:1 of being selected over options adorned with generic seals.

Because the generic seals required practically nothing of the authors or the publishers, did this mean that they provide the promoters who employed them good “bang for the buck”? In our study, they certainly did. This is surprising if one assumes people will evaluate the source of the seal or its implications. It’s less surprising if you assume that even unfamiliar seals will initially have a positive aura and connotation. We should expect such positive influence of generic seals to wane over time and the influence of rigorous seals to increase over time as the audience learns from experience. With exposure and actual interactions, people will become more aware of the meaning of
the seals, the practices of the organizations behind them, and the responses if any to lies and fabrication. Just as your trust in a reliable friend increases over time, positive experiences with effective marks of trustworthiness should produce increased trust and credibility. Thus, we’d expect the relative superiority of TruthSeals to increase over generic seals, as people interact with them more and each becomes more familiar.

Our data make it clear that authors, advertisers and publishers can significantly increase the impact of their messages in contexts where consumers and decision makers are influenced by the credibility of information claims. This will be salutary for authors, sellers, and candidates who have truthful statements to make. The use of such seals to provide ubiquitous meta-data about the credibility of information could usher in an era where everybody can more easily and usefully filter substance from smoke.

References


Appendix 1. Instrument A.

Introduction

Thank you for agreeing to take our survey.
We are a non-profit organization dedicated to improving the ability of citizens and consumers to make good choices.

We will show you a few ads and ask you to rate the advertised entities.

We will also ask you to choose between some competing pairs.

The whole survey should take you less than 15 minutes. We hope you'll take the time required to give us thoughtful responses.

Thanks!

PAGE 2

Product Comparisons

The ads you see can include different types of information from different sources.

We assume you are familiar with newspapers such as the New York Times and the Like button that tallies how many people say they like something.

Another type source you are probably familiar with is consumer ratings, showing some number of stars corresponding to an average of ratings. We are showing BuzzRatings in this case, marked with the following image:

BuzzRating.
Consumer Ratings.

The first products we ask you to rate use only these familiar sources. Below is the first product ad we want you to consider.

<table>
<thead>
<tr>
<th>E-Auto Electric Car</th>
</tr>
</thead>
<tbody>
<tr>
<td>The ad for E-Auto electric car follows:</td>
</tr>
<tr>
<td>Help both your wallet &amp; the environment</td>
</tr>
<tr>
<td>E-Auto electric cars give unlimited mileage</td>
</tr>
<tr>
<td>Easy to use &amp; drive</td>
</tr>
<tr>
<td><strong>Consumer Rating</strong></td>
</tr>
<tr>
<td><strong>E-Auto, approved by UL</strong></td>
</tr>
</tbody>
</table>
E-Auto's ad only cites two sources.

They use Buzz Rating to give an average consumer rating. We don't really know who provided those inputs or how exactly the stars are determined. Buzz Ratings like this will be used throughout the survey.

They cite Underwriters Laboratory as giving their approval, which probably means the car manufacturer told UL that their electronics met UL standards.

1. Based on their ad, rate how appealing you'd find the E-Auto car if you were shopping for an electric. Pick a number between 1 and 7, where 1 is Least Appealing and 7 is Most Appealing.

1 2 3 4 5 6 7 - Most Appealing

Now please look at the following ad for a different electric car.

The Hi-QCar ad also cites Buzz Rating for consumer popularity and UL electronic standards.

This ad introduces another source that you will see several times, the Gold Star Shield, as shown here:

GoldStar
Good company.
The Gold Star Shield is a licensed image. The advertiser pays a fee to be able to affix the image to its statements. Gold Star requires a signed statement from the advertiser as part of its license. The statement says "The advertiser is a good company. It will strive to provide a good product or service consistent with the expectation created by a Gold Star."

Other statements in the ad don't cite any source.

2. Based on their ad, rate how appealing you'd find the Hi-QCar if you were shopping for an electric. Pick a number between 1 and 7, where 1 is Least Appealing and 7 is Most Appealing.

Before moving on, even though you've already rated each car separately, we'd like to know which you'd choose if you had to settle on just one. We'll show the two ads, side by side, and then ask you to indicate which you prefer.

Here are the two ads you've already seen, but repeated side-by-side.

3. If you had to choose one, which of the two would you prefer?
   - E-Auto
   - Hi-QCar

OK, you've now seen the basic types of comparisons we will be asking about. You will see various items to compare based on their ads. You will also see different kinds of information sources cited.

Now we are going to ask you about a couple of different gasolines. The first of these is called Alpha. Here is their ad.
The TruthSeal image is licensed by an independent, non-profit organization that validates claims, makes advertisers guarantee the truth of those claims, collects penalties when claims turn out to be false or misleading, and pays bounties to challengers who present appropriate data falsifying the claims. The advertiser who licenses a TruthSeal mark is guaranteeing the claim is true and accepts independent validation and adjudication of challenges. The seal means "Truth: Guaranteed."

Other statements in this ad don't cite any source.

4. Based on their ad, rate how appealing you'd find Alpha gasoline if you were shopping for fuel. Pick a number between 1 and 7, where 1 is Least Appealing and 7 is Most Appealing.

1 - Least Appealing  2  3  4  5  6  7 - Most Appealing

Beta is an alternative gasoline. Its ad is below. Please look at the ad and then rate the product.
How Seals Increase Ratings and Preferences

The ad for Beta includes many statements from sources you have already seen. No further explanation is required.

5. Please tell us how you’d rate Beta gasoline if you were shopping for fuel. Pick a number between 1 and 7, where 1 is Least Appealing and 7 is Most Appealing.

Before moving on, even though you’ve already rated each gas product separately, we’d like to know which you’d choose if you had to settle on just one.

Here are the two ads you’ve already seen, but repeated side-by-side.

6. If you had to choose one, which of the two would you prefer?

   - Alpha
   - Beta

Now we're going to ask you to compare a couple of athletic shoes. This will be the last "product" comparison. After this, we will change to a different sort of category.
Look at the ad below for Gamma shoes. It includes information and sources you should now be familiar with.

![Gamma Athletic Training Shoes]

The ad for Gamma includes many statements from sources you have already seen. It actually has two statements marked with the Gold Star meaning the advertiser promises the Gold Star licensor to make good products.

7. If you were going to purchase athletic trainers, how would you rate the Gamma shoes, on a scale from 1 to 7?

- 1 - Least Appealing
- 2
- 3
- 4
- 5
- 6
- 7 - Most Appealing

Now please consider the ad for Omega shoes, which includes several statements with different sources and images.

![Omega Athletic Training Shoes]

The ad for Omega includes many statements from sources you have already seen. It actually has two statements marked with the TruthSeal image meaning the independent organization has checked that the claim rests on credible data. The advertiser guarantees the claim is true and pays a penalty if the claim is successfully challenged.
8. Based on its ad, if you were going to purchase athletic trainers, how would you rate the Omega shoes, on a scale from 1 to 7?

- 1 – Least Appealing
- 2
- 3
- 4
- 5
- 6
- 7 – Most Appealing

To wrap up the product comparisons, we'd like to know which athletic training shoes you'd choose if you had to settle on just one.

Here are the two ads you've already seen, but repeated side-by-side.

### Gamma Athletic Training Shoes
**Gamma's ad for its shoes follows:**
- Gamma: Light, Comfortable, Durable
- Gamma trainers are preferred by knowledgeable athletes
- Gamma trainers boost your confidence & increase your capabilities
- Consumer Rating: ★★★★★
- Runner’s World Endorsed

### Omega Athletic Training Shoes
**Omega's ad for its shoes follows:**
- Omega: Nice Feel, Lightweight, Long Lasting
- Omega trainers are preferred to those by Del, Gamma, and Ups.
- Omega trainers rate tops in speed and agility trials
- Consumer Rating: ★★★★★
- Runner’s World Endorsed

9. If you had to choose one, which of the two would you prefer?

- Gamma
- Omega

PAGE 3

**Service comparisons**

We now want you to assess three pairs of competitors in service businesses, using the same basic approach.

The first business is D-Oil, which is described in its advertisement below.
The ad for D-Oil includes many statements and the sources are mostly familiar now. D-Oil also has an endorsement from the Better Business Bureau (BBB).

10. Based on its ad, if you were going to purchase an oil change, how would you rate the D-Oil service, on a scale from 1 to 7?

Serv-U Service competes with D-Oil in the oil change business. Would you please consider the Serv-U ad below and then rate it.

11. Based on its ad, if you were going to purchase an oil change, how would you rate the Serv-U service, on a scale from 1 to 7?

Here are the two ads you've already seen, but repeated side-by-side.
12. If you had to choose one, which of the two would you prefer?

- D-Oil
- Serv-U

Let's now consider a couple of bank service providers, starting with 1st Town's ad below.

**1st Town Banking Services**

**1st Town** ad for service:

- Free checking service (min. bal. required)
- Zero fees at all of our ATMs
- Best customer service in the business

![Consumer Rating](image)

BBB rates 1st Town Excellent

13. Based on its ad, if you were going to sign up for a checking account with a bank, how would you rate 1st Town, on a scale from 1 to 7?

- 1 - Least Appealing
- 2
- 3
- 4
- 5
- 6
- 7 - Most Appealing

Center-1 is an alternative provider, whose ad follows. Please consider it and then rate Center-1.
14. Based on its ad, if you were going to sign up for a checking account with a bank, how would you rate Center-1, on a scale from 1 to 7?

- 1 – Least Appealing
- 2
- 3
- 4
- 5
- 6
- 7 – Most Appealing

Now we’ll repeat both of their ads side by side and ask you to indicate which you’d prefer.

15. If you had to choose one, which of the two would you prefer?

- 1st Town
- Center-1

Let’s wrap up the "Service" category with a couple of dry cleaning establishments. First look at the ad for Joan’s and then rate the service provider.
16. Based on its ad, if you needed dry cleaning, how would you rate Joan’s, on a scale from 1 to 7?

- 1 – Least Appealing
- 2
- 3
- 4
- 5
- 6
- 7 - Most Appealing

Now please consider Mary’s, an alternative dry cleaning service provider. Here is its ad:

17. Based on its ad, if you needed dry cleaning, how would you rate Mary’s, on a scale from 1 to 7?

- 1 – Least Appealing
- 2
- 3
- 4
- 5
- 6
- 7 - Most Appealing

Now we'll repeat both of their ads side by side and ask you to indicate which you'd prefer.
How Seals Increase Ratings and Preferences

18. If you had to choose one, which of the two would you prefer?
   - Joan's
   - Mary's

Comparing Political Candidates

We now want you to assess three pairs of political competitors, using the same basic approach. This will be the last category we ask about.

The first candidate for President in 2012 is Jones. His campaign ad follows:

The ad for Jones includes several statements and the sources are mostly familiar now. Two of the statements are marked with TruthSeal images, and one is a Facebook Like symbol.

The TruthSeal images mean that the claims have been independently validated and are guaranteed to be true. If falsified by a challenger, the challenger wins a bounty and the candidate's campaign would pay a penalty.
19. Based on the ad, if you were going to vote for a Presidential candidate, how would you rate Jones on a scale from 1 to 7?

1 – Least Appealing  2  3  4  5  6  7 - Most Appealing

Smith is opposing Jones in the upcoming election. Would you please consider the ad for Smith which follows.

Smith for US President in 2012
Campaign ad for Smith follows:

Wise, Informed, Experienced
Dedicated to the highest integrity in all responsibilities and at all times
Consistently strives to work across the aisle and govern with bi-partisan support

1,214,109 people Like Smith on Facebook
Rated “qualified” by the NY Times

Smith's ad includes several statements with different sources. One of these sources is new, and it's called the GoodCitizens Seal, as shown below:

GoodCitizens.
Good campaigns

The organization that licenses the GoodCitizens Seal offers the seal for use in political campaigns, so that candidates can apply the seal to sincere statements they make. When using the seal, the candidates and their campaigns commit to conduct good campaigns employing best political practices. The advertiser pays for the right to use the seal in that way.

20. Based on the ad for Smith, if you were going to vote for a Presidential candidate, how would you rate Smith on a scale from 1 to 7?

1 – Least Appealing  2  3  4  5  6  7 - Most Appealing
Here are the two ads you’ve already seen, but repeated side-by-side.

**Jones** for US President in 2012
Campaign ad for **Jones** follows:
- Knowledgeable & Experienced
- Rated “highest integrity” by two independent rating organizations.
- Successfully worked to pass bi-partisan bills repeatedly.
- 1,211,314 people Like Jones on Facebook
- Rated “qualified” by the *NY Times*

**Smith** for US President in 2012
Campaign ad for **Smith** follows:
- Wise, Informed, Experienced
- Dedicated to the highest integrity in all responsibilities and at all times
- Consistently strives to work across the aisle and govern with bi-partisan support
- 1,214,109 people Like Smith on Facebook
- Rated “qualified” by the *NY Times*

21. If you had to choose one, which of the two would you prefer?

- Jones
- Smith

Let’s now consider a couple of candidates for Governor. The first candidate we want you to rate is Able, whose ad appears below.

**Able** for Governor
The campaign ad for **Able** follows:
- Able: Change in 2012, Because We Can
- A successful entrepreneur who knows business
- Informed, experienced, and tested
- 411,314 people Like Able on Facebook
- Able rated “highly qualified” by the *The Times*

22. Based on this ad, if you were going to vote for a candidate for Governor, how would you rate Able, on a scale from 1 to 7?

- 1 – Least Appealing
- 2
- 3
- 4
- 5
- 6
- 7 – Most Appealing

Baker is an alternative candidate, whose ad follows. Please consider the ad and then rate the candidate.
Based on this campaign ad, if you were going to vote for Governor, how would you rate Baker, on a scale from 1 to 7?

1 - Least Appealing
2
3
4
5
6
7 - Most Appealing

Now we'll repeat both of their ads side by side and ask you to indicate which you'd prefer.

If you had to choose one, which of the two would you prefer?

Able

Baker

Let's wrap up the "Political Candidates" category with a couple of politicians running for Senator. First consider the ad for Thompson and then rate that candidate.
25. Based on this ad, if you were voting for Senator, how would you rate Thompson, on a scale from 1 to 7?

1 - Least Appealing
2
3
4
5
6
7 - Most Appealing

Now please consider Williams, an alternative candidate for Senator. Here is the campaign ad:

26. Based on this campaign ad, if you were considering a vote for Senator, how would you rate Williams on a scale from 1 to 7?

1 - Least Appealing
2
3
4
5
6
7 - Most Appealing

Now we'll repeat both of their ads side by side and ask you to indicate which you'd prefer.
27. If you had to choose one, which of the two would you prefer?
   - Thompson
   - Williams

**Last page**

28. What is your age range?
   - Under 18
   - 18-29
   - 30-40
   - Over 50

29. What is your gender?
   - Male
   - Female

30. Do you expect to vote in the 2012 election?
   - Certainly or almost certainly
   - Probably
   - Possible but not likely
   - Certainly not

31. If you have comments or want to make suggestions, please enter them below:

All finished. Click DONE!