How Confidence in Prior Attitudes, Social Tag Popularity, and Source Credibility Shape Confirmation Bias: A Randomized Controlled Web-Based Study about Attitudes towards Antidepressants and Psychotherapy in a Representative German Sample

Mag. Stefan Schweiger 1)
Jun.-Prof. Dr. Aileen Oeberst 1) 2)
Prof. Dr. Ulrike Cress 1) 3)

1) Leibniz-Institut für Wissensmedien, Schleichstr. 6, 72076 Tuebingen, Germany
2) Johannes Gutenberg-Universität Mainz, Germany
3) University of Tuebingen, Schleichstr. 4, 72076 Tuebingen, Germany

Address for correspondence: Mag. Stefan Schweiger, Leibniz-Institut für Wissensmedien, Schleichstr. 6, 72076 Tuebingen, Germany, e-mail: schw.stefan@gmail.com, phone: +43 670 202 3838
How Confidence, Social Tag Popularity, and Source Credibility Shape Confirmation Bias: A Randomized Controlled Web-Based Study about Attitudes towards Antidepressants and Psychotherapy in a Representative German Sample

Abstract

Background: In health-related, web-based information search, people should choose objectively correct information, but they are often misguided by confirmation bias—the tendency to select and evaluate information in line with their prior attitudes. They are also misguided by dubious information, not taking source credibility into account properly.

Objective: We test whether people are prone to confirmation bias in mental health-related information search, particularly (1) if high confidence worsens confirmation bias, (2) if social tags are an appropriate interface to circumvent the influence of prior attitudes, and (3) if people successfully distinguish high and low source credibility. Moreover, we describe attitudes towards the efficacy of the treatment of depression with antidepressants and psychotherapy.

Methods: 520 participants of a representative sample of the German population were recruited on an online platform of a panel company. 250 (48%) completed the fully automated, randomized, controlled web-based study, which was accessible online from November 14th to November 18th 2014, until at least 250 participants completed the survey. Participants provided prior attitudes about antidepressants and psychotherapy. We manipulated (1) confidence by having participants recall situations in which they were confident or doubtful. Next, participants searched for blog posts about the treatment of depression, with social tag clouds differing in (2) tag popularity—either psychotherapy or antidepressant tags were more popular. Finally, we manipulated (3) source credibility with banners indicating high or low expertise of the tagging
community, and we measured tag- and blog post selection, and treatment efficacy ratings after navigation.

**Results:** We observed a tendency to rate psychotherapy (mean = 5.24, SD = 1.10) as more effective than antidepressants (mean = 4.61, SD = 1.19; t(225) = 9.71, P < .001, d = .56.). Tag popularity predicted the proportion of selected antidepressant tags (beta = 0.44, SE = .11, P < .001), and blog posts (beta = 0.46, SE = .11, P < .001). We could not replicate the confidence manipulation (t(224) < 1, P = .78). Participants did not attend to source credibility on banners (t(224) = 1.67, P = .10). When confidence was low (-1 SD), participants selected more blog posts consistent with prior attitudes (beta = -0.26, SE = 0.05, P < .001). Moreover, when confidence was low (-1 SD) and source credibility was high (+1 SD), the efficacy ratings of attitude consistent treatments increased (beta = 0.34, SE = 0.13, P = .01).

**Conclusions:** We found correlational support for defense motivation account underlying confirmation bias in the mental health-related search context. That is, participants did not tend to select objectively correct, but information that supported their prior attitudes.

**Keywords:** Attitude, Psychotherapy, Antidepressants, Culture, Germany, Beliefs, Social Tagging, Tag Cloud, Credibility, Confidence
Introduction

Do people attend to information outside of their filter bubble, and do they distinguish expert from dubious sources on the web? To address these controversial questions [1–3], we investigate confirmation bias, the tendency to favourably select and evaluate attitude consistent information [3–6].

A comprehensive meta-analysis identified two major motivational factors that moderate confirmation bias [7]. First, when we face information that suggests our point of view is wrong, we try to maintain our prior attitudes by choosing and believing attitude consistent information, which is called defense motivation [7–9]. In contrast to this, in some situations we may be genuinely interested in acquiring objectively correct and accurate information [7,8,10]. This accuracy motivation can guide our information search, even when information is not consistent with our prior attitudes [7]. Particularly in the health-context, we should form attitudes independent of our defense mechanisms and base our evaluations on objectively correct information. In the following, we outline three factors that may reduce confirmation bias, given that we are accuracy motivated when searching for mental health-related information.

Confidence and Confirmation Bias

First, low confidence should decrease confirmation bias [7]. However, people tend to be overly confident in prior attitudes and knowledge [11,12] in a large range of domains, such as academic, intellectual, vocational, athletic, and medicine [13]. Overconfidence makes people reluctant to change their attitudes [14], increases confirmation bias [15], and overconfidence increases when having online information at hand, since people mistake having access to information for having knowledge [16].

For the mental health-related context, it is important that confidence varies for people with different mental disorders [17]. For example, individuals who experience anxious and
depressive symptoms, show less than average confidence (but average accuracy) in decision-making tasks [17], which suggests that they could be even less prone to confirmation bias. Looking at how to influence confidence, overconfidence can be reduced when participants reflect on their ability to describe, in a step-by-step manner, the causal functioning of objects to experts [18]. We draw on a manipulation that focussed on people recalling situations where they were either confident or doubtful about their own thoughts (study 3, [14]). When participants recalled situations in which they had been confident (vs. doubtful), and subsequently provided arguments about a controversial topic, they were more (less) confident about their arguments [14,19]. For the current study, one goal was to replicate the manipulation ([14], study 3) with a representative sample, in the mental health context.

We expect that when prior attitudes are held with high confidence, participants preferably select and evaluate attitude consistent information. If participants were defense motivated, high (vs. low) confidence would make them less (vs. more) threatened by attitude inconsistent information, and they would select more attitude inconsistent information, and evaluate it more favourable [7].

**Social Tags as Signposts to Popular Information**

The second influence on confirmation bias occurs when people face cues from socially aggregated information on the web [20–25]. Cues indicating socially aggregated information include star ratings, likes, retweet counts, or social tags. In the case of tagging, tag clouds arise when users label or tag content on the web, such as videos, images or documents (Figure1; [26,27]). When tags from the tagging community are aggregated and presented in tag clouds, the tags represent the consent of a majority of people, and guide information searchers [21,28]. High majority consent or high tag popularity translates into large tags, which attract more attention than smaller tags with less social consent.

We suggest that social tag clouds are particularly non-intrusive and therefore highly suited to circumvent the influence of prior attitudes, since larger tags are visually dominating,
and it has been shown that people who primarily attend to large tags [26,29,30], are more likely to click on large tags [21,31,32], even when large tags are inconsistent with activated associations in memory [31,32], or prior attitudes [21]. Moreover, social consent elicits behaviour that conforms to the majority in offline settings [33,34].

Moreover, people select more trustworthy results when facing a grid-like (vs. list-like) arrangement of search results, similar to social tag clouds [35]. In sum, tag clouds should be suited to decrease the influence of prior attitudes in information search and reduce confirmation bias.

**Figure 1.** The tag clouds used in the present study. Either psychotherapy (left), or antidepressants (right) were more popular.

### Source Credibility of the Tagging Community

People do not always successfully consider high quality information [36,37], particularly not when browsing user generated content [36]. Two meta analyses concluded that personal characteristics [37], as well as platform characteristics [36] play an important role. The relationship between manipulated source credibility and perceived information credibility is higher for student samples (vs. non-student samples) [37], and user generated content that is presented on common websites (vs. blogs and discussion boards) [36].

Besides education, epistemic beliefs can influence how people perceive source credibility. For example, when searching information on two competing therapies for Bechterew’s disease, not all participants took source credibility into account [38]. Particularly participants who viewed
the web as a reliable (vs. not reliable) source of accurate knowledge did not reflect upon source credibility, they viewed URLs for a shorter time, and selected less search results at the bottom of the page.

To our knowledge, there is a single study using tag clouds to investigate source credibility in the health context [39]. People searched for information on how to manage diabetes on a health forum with tag clouds [39]. In the first task, they searched for information that was of general interest, and in a second task, they searched for credible information. The tag cloud consisted of 12 filler tags (e.g. glucose, diet, exercise), and six tags indicating source credibility of content (author, date, quote, reference, statistics, testimonial). When participants browsed for general interest, only one third used at least a source credibility tag. When explicitly asked to take source credibility into account, 90% used at least one source credibility tag. It remains an open question, whether people in a representative sample take the source credibility on a social tagging platform into account. In line with the accuracy motivation account, we expect that if information searchers recognize high source credibility, they will select more tags and related blog posts in total, regardless of whether attitude consistent or attitude inconsistent tags are more popular in the social tag cloud. If, on the other hand, people showed defense motivation [7], they would avoid attitude-inconsistent tags and blog posts with high source credibility, and evaluate it less favourable.

**Prior Attitudes towards Antidepressants and Psychotherapy**

With respect to the treatment of depressive disorders, people clearly favour psychotherapy over antidepressants [40–46]. Attitudes of laypeople manifest in estimated treatment efficacy as well as treatment recommendations [21,43,45,47,48]. People believe antidepressants to be little to moderately effective, whereas psychotherapy is believed to be moderately to highly effective [21]. Since literature shows about equal, moderate efficacy of both types of treatment [49–51], people’s attitudes and recommendations are biased.
We expect more positive prior attitudes towards psychotherapy than towards antidepressants in the German population, and with the current study we aim to describe the magnitude of the psychotherapy preference and present the arguments which shape these biased attitudes.

**Hypotheses**

First, we expect that people’s attitudes (H1a) and efficacy ratings (H1b) prior to navigation are more favourable for psychotherapy than for antidepressants.

We expect that high (vs. low) *confidence* leads to a more pronounced confirmation bias, an increased selection of attitude consistent tags (H2b) and blog posts (H2c), and this will strengthen the attitudes people already had before navigation (H2d). So, when prior attitudes favour psychotherapy, and confidence is high, participants prefer psychotherapy tags and blog posts, and change their attitudes even more towards psychotherapy. If confidence is low, prior attitudes should not be related to selection of tags and blog posts, and attitude change.

*Tag popularity* should circumvent the influence of prior attitudes, so participants select popular tags more frequently than less popular tags (H3a) and blog posts (H3b). Consequently, attitudes change in line with tag popularity (H3c).

Participants distinguish high from low *source credibility* (H4a). When tags and blog posts were collected by experts (vs. novices), participants click on more tags (H4b), and blog posts (H4c) overall, independent their prior attitudes, and people should show more attitude change for both treatments (H4d).

**Methods**

**Participants**

A representative sample with respect to age and gender was randomly drawn from a pool of a panel company. 520 (100%) participants started the fully automated online study, 250 (48%) completed it, 7 (1%) withdrew their data, further 17 (3%) participants were dropped as they did
not provide responses (Figure 2). Age of the remaining 226 (43%) participants ranged from 18 to
60 years (mean = 40.36, SD = 12.17), 113 (50%) were female (Table 1). With respect to
familiarity of the technology used in the study, 56 (25%) stated they were familiar with the term
“tag cloud”, 83 (37%) stated they had already clicked on single tags to navigate the web. Ethical
approval was granted by the Ethical Committee of the Knowledge Media Research Center (LEK
2014/006).
Table 1. Sample characteristics.

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>N</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Education</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Qualified job</td>
<td>18</td>
<td>8.0</td>
</tr>
<tr>
<td>Abitur certificate</td>
<td>102</td>
<td>45.1</td>
</tr>
<tr>
<td>University degree</td>
<td>53</td>
<td>23.5</td>
</tr>
<tr>
<td>Other</td>
<td>53</td>
<td>23.5</td>
</tr>
<tr>
<td><strong>Age</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18-19</td>
<td>9</td>
<td>4.0</td>
</tr>
<tr>
<td>20-29</td>
<td>46</td>
<td>20.4</td>
</tr>
<tr>
<td>30-39</td>
<td>45</td>
<td>19.9</td>
</tr>
<tr>
<td>40-49</td>
<td>65</td>
<td>28.8</td>
</tr>
<tr>
<td>50-59</td>
<td>52</td>
<td>23.0</td>
</tr>
<tr>
<td>60</td>
<td>9</td>
<td>4.0</td>
</tr>
<tr>
<td><strong>Gender</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>113</td>
<td>50.0</td>
</tr>
<tr>
<td>Female</td>
<td>113</td>
<td>50.0</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>226</td>
<td>100</td>
</tr>
</tbody>
</table>

Figure 2. Participant flow diagram.

Procedure and Design

The study comprised a 2 (confidence: high, low) x 2 (tag popularity: antidepressants high, psychotherapy high) x 2 (tagging source credibility: high, low) between-subjects design.

Participants enrolled via online portal of a private panel company, which linked to our survey, and participants were offered 4 € to complete it. First, participants were welcomed and informed
that they could withdraw participation at any point. Participants were granted anonymity and asked to provide informed consent by clicking the button to start the study, after which they were randomly assigned to one of the six experimental conditions by a computerized random number procedure. Then, for prior attitudes, we asked participants to state pro and contra arguments regarding antidepressants and psychotherapy (Pretest tasks I, Figure 3). Next, they rated the efficacy of antidepressants and psychotherapy on scales. Then they provided responses for an allegedly unrelated pilot study, which served to manipulate confidence [14]. Participants were asked to recall situations in which they had either felt confident or doubtful about their own knowledge (study 3, [14]). After this, they were asked to think back to their arguments regarding psychotherapy and antidepressants and they rated how confident they were about the arguments they had provided before. This rating served as a manipulation check for confidence.

**Figure 3.** Experimental procedure.

Next, participants searched for information about treatment efficacy to provide treatment advice for a hypothetical, closely related person. To manipulate source credibility, we informed them that the blog post and responding tag had been gathered by a community of either experts in the field such as experienced psychiatrists and by psychotherapists (high source credibility condition), or by psychology students and medical students in their first semester (low source credibility condition). To manipulate tag popularity, either psychotherapy or antidepressant tags
were larger (Figure 1). They could also provide tags for blog posts. After five minutes of browsing in the tagging environment, a “Next” button appeared and from then on participants could decide, when to stop browsing tags and related blog posts. After navigation, participants rated source credibility (manipulation check) and provided efficacy ratings again. At the end of the study participants could provide feedback in a text box.

Materials

For the information search task, we provided a tagging environment (Figure 4). We presented 14 tags on the right side of the screen. Five tags represented psychotherapy and five tags represented antidepressant treatments. Four filler tags were irrelevant for treatment (prejudice, media coverage, societal relevance, prevalence). Depending on the experimental condition either psychotherapy-related tags or antidepressants-related tags were larger (i.e., more popular). When participants clicked on a tag, three blog posts were presented to the left (Figure 4). Each blog post described a symptom of depressive disorders and scientific studies on the efficacy of the treatment. In a pilot study, we had assured that the blog posts had equal persuasiveness. First only the headline and the first sentence of each of the three related blog posts was shown. To read the full post, participants clicked on “(more…)”. 
Independent Variables

Prior Attitudes

As index of pro and contra arguments for psychotherapy and antidepressants, we subtracted the number of arguments favouring antidepressants (contra psychotherapy, pro antidepressants) from the number of arguments favouring psychotherapy (pro psychotherapy, contra antidepressants). Positive values of this index thus indicate a preference for psychotherapy. Arguments were rated by two raters ($r = .78$, $P < .001$), differences were resolved by agreement.

Confidence

We adapted the experimental procedure by Petty and colleagues (study 3, [14]), and participants recalled situations in which they had felt confident or doubtful about their own knowledge, using five input text boxes for five minutes.

Tag Popularity
For the psychotherapy popular group, psychotherapy tags were larger, and for the antidepressant popular group, antidepressants tags were larger (Figure 1).

**Source Credibility**

On top of the page, banners showed that either alleged college students (low source credibility; Figure 5) or domain experts (high source credibility; Figure 6) had collected and tagged the blog posts. After the search task, participants rated the source credibility of the information on a scale from 1 (not at all) to 7 (highly).

**Confidence Ratings (Manipulation check)**

After participants listed situations in which they had been (un)confident, they rated confidence in their own arguments regarding prior attitudes on a scale from 1 (not at all) to 7 (highly). They were asked how the following words described their arguments: obvious, dubious, justified, credible, factual, well-founded, persuasive and objective (Cronbach alpha = .88).

**Source Credibility Ratings (Manipulation check)**

Participants rated the degree to which the following words described the tagging community: informed and competent (r = .70, P < .001).

**Figure 5.** Banner for the low source credibility condition.

**Figure 6.** Banner for the high source credibility condition.

**Dependent Variables**

**Efficacy Ratings (Attitude Change).** Participants agreed to statements on the efficacy of psychotherapy and antidepressants on a scale from 1 (completely disagree) to 7 (completely agree).
agree), before (antidepressants Cronbach alpha = .89, psychotherapy Cronbach alpha = .92), and after navigation (antidepressants Cronbach alpha = .94, psychotherapy Cronbach alpha = .95). To predict attitude change with respect to treatment preference, we derived a difference index score, subtracting the antidepressant from psychotherapy treatment ratings.

Besides attitude change in terms of treatment preference, we analysed pooled attitude change by taking the sum of efficacy ratings for both treatments prior to and after navigation (divided it by the number of items for interpretability).

**Tag and Blog Post Selection.** To measure attitude consistent navigation, we recorded the number of tags and blog posts selected for each treatment category (0 = psychotherapy, 1 = antidepressants).

## Results

All analyses presented were conducted with the R software (Version 3.3.4), raw data and the analysis script can be found in the Multimedia Appendix 1.

**Prior Attitudes**

As expected in H1a we found that participant’s prior attitudes favour psychotherapy over antidepressants. Participants stated more arguments for psychotherapy (mean = 1.69, SD = 1.77) than for antidepressants (mean = 1.06, SD = 1.51; t(225) = 5.30, P < .001, d = .26), and they stated more arguments against antidepressants (mean = 1.51, SD = 1.53) than against psychotherapy (mean = 0.73, SD = 1.54; t(225) = 8.13, P < .001, d = .34). We also descriptively analysed arguments and pooled them into qualitative categories (Figure 7).

**Figure 7.** Arguments for and against the two treatments.
With H1b we expected that people would provide more favourable efficacy ratings for psychotherapy compared to antidepressants before navigation. Participants rated statements about the efficacy of both treatments on 8 items, on a scale from 1 to 7 (Figure 8). Since the internal consistency was high for both scales (antidepressants Cronbach alpha = .89, psychotherapy Cronbach alpha = .92), we pooled them. A paired $t$-Test showed a moderate effect on the preference for psychotherapy (mean = 5.24, SD = 1.10) over antidepressants (mean = 4.61, SD = 1.19; $t(225) = 9.71$, $P < .001$, $d = .56$; see items and response distribution in Figure 8). In sum, prior attitudes measured via pro and contra arguments, as well as via efficacy ratings, favoured psychotherapy over antidepressants. Both measures were moderately correlated ($r = .41$, $P < .001$).

**Figure 8.** Prior attitudes about psychotherapy and antidepressants prior to information search.
Manipulation Checks

Confidence

Contrary to our expectations (H2a), we could not replicate the confidence manipulation (study 3, [14]). After recalling situations in which they had been confident (mean = 4.64, SD = 1.20), participants were not more confident about their arguments, compared to recalling situations in which they had been doubtful (mean = 4.68, SD = 1.09; t(224) < 1, P = .78).

Source Credibility

In contrast to our expectation (H4a), source credibility ratings in the high source credibility condition (mean = 4.87, SD = 1.26) did not significantly differ from source credibility ratings in the low source credibility condition (mean = 5.16, SD = 1.32; t(224) = 1.67, P = .10).

Since the confidence and source credibility manipulations were ineffective, we used respective manipulation check scores in the following regression analyses as predictors.

Confidence in Prior Attitudes

Tag Selection

To analyse attitude consistent tag selection, we conducted logistic regressions with the dependent variable clicks on attitude (in)consistent treatment tags. The number of clicks on the
respective treatment (0 = psychotherapy tag selected, 1 = antidepressant tag selected) was entered in a logistic regression (Table 2). As predictors we entered prior attitudes and tag popularity (0 = Psychotherapy tags popular, 1 = antidepressant tags popular), confidence ratings and source credibility ratings (see independent variables). We included 2-way interaction terms (Table 2, Step 2), and tested for the presence interactions with likelihood ratio tests [52,53].

We expected that high confidence should strengthen the relationship between prior attitudes and the proportion of clicks on attitude consistent tags (H2b). However, there was no significant interaction of the predictors confidence in prior attitude ratings and prior attitudes, predicting the selection of antidepressant tags (Table 2, Step 2). As likelihood ratio tests showed, including three-way interaction ($\chi^2(10) = 4.91, P = .90$), and four-way interaction ($\chi^2(11) = 4.98, P = .93$), did not improve model fit.

**Table 2. Selection ratio of antidepressant tags.**

<table>
<thead>
<tr>
<th></th>
<th>Step 1</th>
<th>Step 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>B</td>
<td>SE</td>
</tr>
<tr>
<td>Intercept</td>
<td>-0.39</td>
<td>0.08</td>
</tr>
<tr>
<td>Prior attitudes</td>
<td>0.03</td>
<td>.37</td>
</tr>
<tr>
<td>Confidence score</td>
<td>0.002</td>
<td>0.05</td>
</tr>
<tr>
<td>Tag popularity</td>
<td>0.44</td>
<td>0.11</td>
</tr>
<tr>
<td>Source credibility score</td>
<td>-0.005</td>
<td>0.04</td>
</tr>
<tr>
<td>PA x confidence score</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Model fit</td>
<td>$\chi^2(4) = 17.89, P = .001$</td>
<td>Model fit change (vs. Step 1): $\chi^2(1) = 0.20, P = .65$</td>
</tr>
</tbody>
</table>

B, Beta coefficients, continuous predictors were centered; SE, Standard Error of Beta coefficients; PA, Prior attitudes.

**Blog Post Selection**

A second logistic regression used the same predictors as in the regression predicting tag selection but with blog post selection as criterion variable (Table 3). We expected that high confidence should strengthen the impact of prior attitudes and consequently lead to increased proportion of clicks on attitude consistent blog posts (H2c). We observed an interaction between
confidence and prior attitudes (beta = 0.11, SE = .02, P < .001). To disentangle the interaction, we compared slopes for high (+1 SD) and low (-1 SD) confidence ratings. This showed that when confidence ratings were low (-1 SD), participants selected a higher proportion of blog posts that were in line with their prior attitudes (beta = -0.26, SE = 0.05, P < .001; Figure 9). When confidence ratings were high (+ 1SD), there was no association with prior attitudes (beta = 0.02, SE = .03, P = .57; Figure 9). In contrast to our expectation, and in line with the defense motivation account, when confidence was low but not high, there was an association between prior attitudes and selection of attitude consistent blog posts.

Table 3. Selection ratio of antidepressant blog posts.

<table>
<thead>
<tr>
<th></th>
<th>Step 1</th>
<th>Step 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>B</td>
<td>SE</td>
</tr>
<tr>
<td>Intercept</td>
<td>-0.75</td>
<td>0.07</td>
</tr>
<tr>
<td>Prior attitudes</td>
<td>-0.05</td>
<td>0.03</td>
</tr>
<tr>
<td>Confidence score</td>
<td>-0.11</td>
<td>0.05</td>
</tr>
<tr>
<td>Tag popularity</td>
<td>0.44</td>
<td>0.11</td>
</tr>
<tr>
<td>Source credibility score</td>
<td>0.02</td>
<td>0.04</td>
</tr>
<tr>
<td>PA x confidence score</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Model fit:

χ²(4) = 30.41, P < .001
χ²(1) = 25.56, P < .001

B, Beta coefficients, continuous predictors were centered; SE, Standard Error of Beta coefficients; PA, Prior attitudes.

Figure 9. Predicted proportion of antidepressant blog posts selected, for high (+1 SD) and low (-1 SD) confidence (95% CI), with negative values indicating a preference for AD over PT.
Compared to the model including the 2-way interaction term (Table 3, Step 2), neither including three-way interaction term ($\chi^2(4) = 5.82$, $P = .21$), nor including the four-way interaction term ($\chi^2(5) = 5.99$, $P = .31$) yielded a better model fit (all respective lower order interaction terms were included as well).

**Attitude Change**

We conducted multiple linear regressions. First, with the predictor variables prior attitudes, confidence ratings, and source credibility ratings (all centered), and the dichotomous variable tag popularity ($0 =$ psychotherapy popular, $1 =$ antidepressants popular). Additionally we included the predictor difference score of efficacy ratings (antidepressants subtracted from psychotherapy) prior to navigation, to analyze attitude change with a covariate approach [54]. As criterion for attitudes after navigation, we included the difference score of efficacy ratings (Table 4).

We expected that high (vs. low) confidence would lead to higher confirmation bias and decreased attitude change, so for people who hold their attitudes with high (vs. low) confidence, prior attitudes should be more strongly associated with attitudes after navigation (H2d). The expected interaction between confidence and prior attitudes was not significant (Table 4, Step 2). However, the association between confidence and prior attitudes depended on source credibility (Table 4, Step 3). To disentangle this 3-way interaction, simple slopes were tested on low (-1 SD) and high (+1 SD) levels of source credibility ratings and confidence ratings. This revealed a
strong association between prior attitudes and treatment efficacy ratings after navigation, for participants with lower ratings of confidence (-1 SD) and high source credibility ratings (+1 SD; beta = 0.34, SE = 0.13, P = .01), but no association for high confidence ratings (+1 SD) and low source credibility ratings (-1 SD; beta = 0.11, SE = 0.06, P = .053). There was also no association with low confidence (-1 SD) and low source credibility ratings (-1 SD; beta = 0.08, SE = 0.10, P = .42), and with high confidence (+1 SD) and high source credibility ratings (+1 SD; beta = -0.04, SE = .06, P = .50; Figure 10).

Table 4. Treatment Efficacy Ratings (AD-PT) after Navigation.

<table>
<thead>
<tr>
<th></th>
<th>Step 1</th>
<th></th>
<th>Step 2</th>
<th></th>
<th>Step 3</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>B</td>
<td>SE</td>
<td>P</td>
<td>B</td>
<td>SE</td>
<td>P</td>
</tr>
<tr>
<td>Intercept</td>
<td>0.70</td>
<td>0.08</td>
<td>&lt;.001</td>
<td>0.73</td>
<td>0.08</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Efficacy ratings (PT-</td>
<td>0.79</td>
<td>0.06</td>
<td>&lt;.001</td>
<td>0.78</td>
<td>0.06</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>AD) prior to navigation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tag popularity</td>
<td>-0.05</td>
<td>0.12</td>
<td>.64</td>
<td>-0.06</td>
<td>0.12</td>
<td>.60</td>
</tr>
<tr>
<td>PA</td>
<td>0.08</td>
<td>0.03</td>
<td>.01</td>
<td>0.10</td>
<td>0.06</td>
<td>.005</td>
</tr>
<tr>
<td>Confidence score</td>
<td>-0.08</td>
<td>0.05</td>
<td>.16</td>
<td>-.10</td>
<td>0.06</td>
<td>.08</td>
</tr>
<tr>
<td>Source credibility score</td>
<td>0.04</td>
<td>0.05</td>
<td>.34</td>
<td>0.05</td>
<td>0.05</td>
<td>.33</td>
</tr>
<tr>
<td>PA x confidence score</td>
<td></td>
<td></td>
<td></td>
<td>-0.04</td>
<td>0.03</td>
<td>.16</td>
</tr>
<tr>
<td>PA x source credibility score</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Confidence score x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>source credibility score</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PA x confidence score x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>source credibility score</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Model fit: adj. R² = .49, F(5,220) = 43.68, P < .001
Model fit (vs. Step 1): Δ adj. R² < .01, F(1,219) = 1.94, P = .16
Model fit (vs. Step 2): Δ adj. R² = .02, F(3,216) = 3.88, P = .01

SE, Standard Error of Beta coefficients; PA, Prior attitudes.
B, Beta coefficients, the continuous predictors were centered;

Figure 10. Prior attitudes, confidence and source credibility and treatment efficacy ratings after navigation, with negative values on all axes indicating a preference for AD over PT.
Tag Popularity of Treatments in the Social Tag Cloud

Tag Selection

In H3a we expected that popular treatment tags would be selected more often, independent of prior attitudes. To test this, we used a logistic regression model as described in the previous confidence section on tag selection (see Table 2, step 1). Tag popularity was the only significant predictor for the proportion of selected antidepressant tags (beta = 0.44, SE = .11, P < .001). This supports H3a, as participants selected a larger proportion of popular tags in the tag cloud. They did this independent of their prior attitudes as prior attitudes were not associated with tag selection.

Blog Post Selection

We also expected that participants would select more blog posts when related tags were more popular (H3b). We tested this with the logistic regression as described in the previous confidence section on blog post selection (see Table 3). This showed that participants selected a larger proportion of blog posts when related tags were popular in the tag cloud, supporting H3b (beta = 0.44, SE = .11, P < .001; Table 3, Step 1).

Attitude Change
We expected in H3c that the attitude change of would depend on tag popularity. More popular tags for a treatment should be associated with higher ratings of treatment efficacy. We conducted a linear regression analysis, as described in the previous confidence section (see also Table 4). We did not find an influence of tag popularity on efficacy ratings after navigation (beta = -0.05, SE = .12, P = .64; see Table 4, Step 1).

**Source Credibility of the Tagging Community**

**Tag Selection**

We expected that when the tagging community consists of experts (vs. novices), participants click on more tags (H4b). To test this, we conducted a negative binomial regression with the continuous, centered predictors source credibility score, prior attitudes, confidence score, the dichotomous predictor tag popularity (0 = psychotherapy tags popular, 1 = antidepressant tags popular), and the dependent variable number of selected tags (Table 5). We did not find support for H4b, as the total number of selected tags was not associated with source credibility ratings.

**Blog Post Selection**

We expected that when the tagging community consists of experts (vs. novices), participants click on more blog posts (H4c). We conducted another negative binomial regression with the continuous, centered predictors source credibility score, prior attitudes, confidence score, the dichotomous predictor tag popularity (0 = psychotherapy tags popular, 1 = antidepressant tags popular), and the dependent variable total number of selected blog posts (Table 5). As with the number of selected tags, perceived source credibility did not predict the total number of selected blog posts, not supporting H4c.

**Table 5. Tags and blog posts selected.**
### Table 6. Treatment Efficacy Ratings (AD+PT) after Navigation.

<table>
<thead>
<tr>
<th>Ratings T2 (AD+PT)</th>
<th>B</th>
<th>SE</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>10.26</td>
<td>.12</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Efficacy ratings prior to navigation (AD + PT)</td>
<td>0.76</td>
<td>.05</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Tag popularity</td>
<td>0.02</td>
<td>.18</td>
<td>.93</td>
</tr>
<tr>
<td>PA</td>
<td>0.08</td>
<td>.04</td>
<td>.06</td>
</tr>
<tr>
<td>Confidence in PA score</td>
<td>0.14</td>
<td>.09</td>
<td>.12</td>
</tr>
<tr>
<td>Source credibility score</td>
<td>0.24</td>
<td>.07</td>
<td>&lt;.001</td>
</tr>
</tbody>
</table>

Model fit:
adj. $R^2 = .16$, $F(5,220) = 9.47$, $P < .001$

B, Beta coefficients; SE, Standard Error of Beta coefficients; continuous predictors and criterion were centered; PA, Prior attitudes.
Discussion

With this randomized, controlled study, we aimed to investigate prior attitudes about antidepressants and psychotherapy, and the tendency to confirm prior attitudes when selecting and evaluating mental health-related information. We presented three factors to counter confirmation bias: popularity of treatment tags in a social tag cloud, confidence and the source credibility of the tagging community. We expected that people would select and favourably evaluate attitude inconsistent content, when confidence was low (vs. high). And we expected that source credibility and tag popularity should influence selection of tags independent of prior attitudes. We could not replicate the confidence manipulation ([14], study 3), and participants did not distinguish source credibility as presented by banners, therefore we used manipulation check scores for correlational analyses.

As expected people in the German population rated psychotherapy as more effective than antidepressants, and they reported according beliefs. Increasing tag popularity increased selection of tags, independent of prior attitudes and confidence. In contrast to our expectations, higher source credibility was not associated with increased tag or blog post selection. Participants with high confidence were more open to select attitude inconsistent blog posts, which is in line with the defense motivation account, but not with the accuracy motivation account [7]. Moreover, we found people with low confidence rated treatment efficacy in accordance with their prior attitudes, but only when perceived source credibility was high.

Social Tags to Reduce Confirmation Bias

We expected that social tag clouds are a non-intrusive interface to circumvent prior attitudes, and popular tags would be selected more often in dependent of prior attitudes. We found that people selected popular tags and related blog posts more often. We think that these
findings highlight the important role of popular content on the web, also in the context of mental health-related selection of information. When two treatment options are presented to a searcher, searchers will be guided by more popular information, even independent of their prior attitudes. This could help to design online platforms in which it is desirable to minimize the influence of prior attitudes and maximize the influence of a community. However, in our study we could not find that the effect of tag popularity transfers to attitude change. It seems that the biased attitudes towards psychotherapy are too robust to change in the light of attitude inconsistent tags.

Defense Motivation in Mental Health-Related Information Search

We expected that people would be guided by accuracy motivation when searching mental health-related information. People should strive to select and evaluate information that is objectively correct, regardless of their prior attitudes. In contrast this, the pattern of results suggests that information searchers were defense motivated, which is the tendency to confirm prior attitudes, in order to avoid dissonant cognitions and to maintain a positive view of oneself [7,10].

This was reflected in blog post selection and resulting attitude change. We found that low confidence was associated with selecting attitude consistent blog posts, which suggests that participants may have felt increased threat under low confidence.

The findings on attitude change provide further support for the defense motivation account. People with high confidence were expected to change their attitudes in line with their prior attitudes. However, again, we found the opposite pattern. When confidence was low, not high, people’s attitudes after navigation were polarized in line with their prior attitudes. In contrast to blog post selection, however, this pattern was only found when source credibility was high but not when source credibility was low. This suggests that attitude inconsistent information could have posed a double threat when source credibility was high, in combination with low
confidence. In all other instances, there was no association between prior attitudes and attitude change.

Taken together, the pattern of results resonates with the defense motivation account [7], but not with the accuracy motivation account as we had expected. Future studies should test whether this result extends to other health-related domains, beyond treatment of depression, and to other information platforms as well.

**Source credibility and Confirmation Bias**

People do sometimes recognize source credibility on the web [36,35,56], however, participants did not rate practitioners with years of experience as a more credible, compared to students of health-related subjects in their first semester. One possible explanation for this is that the banners on top of the page were too subtle.

Moreover, for non-student samples (vs. student samples) [37], and user generated content that is presented on common websites (vs. specific platforms) [36], the relationship between manipulated source credibility and perceived information credibility is higher. This might explain that for this representative sample on a specific tagging platform people did not distinguish high from low source credibility.

**Public Attitudes towards Antidepressants and Psychotherapy**

As for student [21] and representative samples in Germany [40], we also expected prior attitudes to be more positive for psychotherapy than for antidepressants, and we found an according moderate effect. The results about the specific beliefs show that people are not satisfied with the current accessibility of mental health care services, and the German population seems to have poor mental health literacy when it comes to side effects of antidepressants. Side effects that can be found in the literature, such as nausea, insomnia, somnolence, fatigue, sexual dysfunction, and weight gain [57,58], were rarely associated with antidepressants.
Limitations

The current study suggests that results for confidence and its interplay with source credibility are in line with predictions of defense motivation, however, due to the correlational design, potential correlated confounding influences could be at work and potentially could have been overlooked.

Moreover, all blog posts highlighted the efficacy aspect of prior attitudes, other important issues such as side effects or treatment of psychological causes were not mentioned in the blog posts. Thus, only one aspect related to prior attitudes, namely treatment efficacy was addressed in the blog posts. Finally, all blog posts were formulated positively, such that information revealing limitations and boundary conditions of the treatments were addressed in the blog posts.

Acknowledgements

This study was funded by the Knowledge Media Research Center. We thank André Klemke and Manfred Knobloch who designed the tagging platform.

Conflicts of Interest

None declared.

Abbreviations

AD = Antidepressants
PT = Psychotherapy
PA = Prior attitudes

Multimedia Appendix 1

Original data and R analysis file.

References

1. Liao QV, Fu W-T. Beyond the filter bubble. Proc SIGCHI Conf Hum Factors Comput Syst


36. Ma T (Jennifer), Atkin D. User generated content and credibility evaluation of online


44. Jorm AF. Mental health literacy: Public knowledge and beliefs about mental disorders. Br


51. Amick HR, Gartlehner G, Gaynes BN, Forneris C, Asher GN, Morgan LC, Coker-


54. Van Breukelen GJP. ANCOVA versus change from baseline had more power in randomized studies and more bias in nonrandomized studies. J Clin Epidemiol 2006;59(9):920–925. PMID:16895814


