

Less Than I Expected and oh so True? On the Interplay Between Expectations and Framing Effects in Judgments of Truth

Mariela E. Jaffé¹ & Rainer Greifeneder¹

Abstract

This manuscript investigates the negativity bias in truth judgments, which holds that negatively (versus positively) framed statements are more likely to be judged true. Throughout four studies we find that expectations moderate the negativity bias. In particular, Study 1 failed to replicate the negativity bias with standard items. In Study 2 we investigated individuals' expectations regarding the statements. When systematically adjusting the percentages in negatively framed statements to be lower than expected, a negativity bias occurred in Study 3. Building on this knowledge in Study 4, we systematically decreased and increased percentages in both framing conditions, investigating the impact of under- versus overestimation. While expectations had no consistent effects for positive frames, overestimation (versus underestimation) led to a higher likelihood of perceived truth in the negative framing condition. Results are discussed in context of current research on language and social psychology as well as post truth politics.

Keywords

framing effects, communication style, information processing, cognition, judgments of truth

¹University of Basel, Basel, Switzerland

Corresponding Author:

Mariela E. Jaffé, Center for Social Psychology, University of Basel, Basel, Switzerland
Email: mariela.jaffe@unibas.ch

On April 17, 2018, Donald Trump, then president of the United States of America, tweeted the following: “Rasmussen just came out at 51% Approval despite the Fake News Media [...]” (RealDonaldTrump, 2018). This tweet referred to the Rasmussen reports that assessed the approval rate in regards to the president’s performance. How did individuals evaluate this piece of information? Did they perceive this piece of information as true? And would they have perceived it differently if the same notion was framed negatively, focusing on a 49% *disapproval* rate? Recent findings suggest that the frame of a statement, meaning whether it focuses on something conceptually positive versus negative, plays an important role in the evaluation of statistical statements. In particular, Hilbig (2009, 2012a, b) showed that content-wise identical statements are more likely to be deemed true when framed negatively (e.g., focusing on the disapproval rate) compared to positively (e.g., focusing on the approval rate), which has been coined the *negativity bias in judgments of truth*. According to these findings, framing Donald Trump’s tweet in terms of disapproval (compared to approval) should lead to a higher probability of being judged true. Here we extend previous findings and argue that the negative frame is more credible in some conditions than in others. More specifically, we empirically demonstrate that individuals’ expectations regarding the to-be-evaluated issues (e.g., what are individuals’ expectations in regards to the (dis)approval rates) play a crucial role in whether a negatively or a positively framed statement is more likely to be judged true. Especially when negatively framed messages are less negative than expected, judgments are more likely to lean in the direction of “true” compared to “false,” for a good reason, as we argue later.

Truthfulness is one of the cornerstones of our society. Research suggests that individuals generally show a willingness to believe new information to be true (truth bias, e.g., Burgoon, Blair, & Strom, 2008), especially when the information is presented by one’s in-group (Clementson, 2018). However, the debate about fake news casts doubt on whether a general truth bias is adaptive, as apparently politicians, managers, and the press do not necessarily speak

the (entire) truth. In light of this trend, the Oxford Dictionary chose post-truth as word of the year 2016, an adjective described as “relating or denoting circumstances in which objective facts are less influential in shaping public opinion than appeals to emotion and personal beliefs” (Oxford Dictionary, 2016). Although the concept of post-truth has existed for the past decade, Oxford Dictionary argued that the frequency of usage spiked in the context of the EU referendum in the United Kingdom and the 2016 presidential election in the United States and has therefore been associated with the particular noun of post-truth politics. The notion of post-truth reflects the general assumption that we are entering a time “in which the specified concept [of truth] has become unimportant or irrelevant,” especially in the context of politics. In such a world of uncertainty, alternative facts, and post-truth politics, not simply believing statements to be true but distinguishing between what is true or false becomes more and more important – and for scientists understanding the mechanisms of when individuals are more or less likely to evaluate something as true is of special interest.

How do Individuals Judge Truth?

Previous research has identified different mechanisms that influence whether individuals judge a statement as true or false. Particularly influential aspects are: To what extent statements are coherent with individuals’ beliefs (Dunwoody, 2009); whether or not individuals succeed in finding matching information in their memory (Reder, 1982); to what extent the information is consistent and plausible (Reder, 1982); and how easily (fluently) individuals can process the information (Dechêne, Stahl, Hansen, & Wanke, 2010; Reber & Schwarz, 1999). Furthermore, the valence framing of the statement has been identified as influential, indicating that individuals are more likely to evaluate a content-wise identical statement as true when framed negatively compared to positively (Hilbig, 2009, 2012a, b), as further detailed next.

The Negativity Bias in Judgments of Truth

Hilbig (2009) first reported that formally equivalent messages are more likely to be deemed true when framed negatively compared to positively. In a typical experiment (Hilbig, 2012a), participants are provided with a number of statements, which are framed either negatively (“In the Dominican Republic, 30% of the population are not reached by vaccination measures”) or positively (“In the Dominican Republic, 70% of the population are reached by vaccination measures”). Importantly, the valence of the frame refers only to the topics or attributes that the sentence is focusing on (such as approval/disapproval rate in the introductory example) and therefore conceptualized as *concept frame*. Within this conceptualization, the question is whether the sentences’ overall framing refers (normatively) to something good (approval rate or being reached by vaccination measures) or something bad (disapproval rate or not being reached by vaccination measures).

This conceptualization, however, does not take into account the *valence of the concept* in question (the sentence’s target concept, e.g., satisfaction as concept with positive valence versus dissatisfaction as concept with negative valence) nor the *negation* (the semantic operation of negating a concept). Yet, concept valence and negation often work in tandem to produce concept framing. That is, “not satisfied” (*negated positive concept valence*) or “dissatisfied” (*negative concept valence*) both result in a negative frame; “not dissatisfied” (*negated negative concept valence*) or “satisfied” (*positive concept valence*) both result in a positive frame. This conceptualization also does not consider the *subjectively construed consequentiality*: Even a negative frame may be something positive, if it is less prevalent than expected. From the perspective of the original literature, however, a sentence focusing on divorce will be considered as negatively framed, and a sentence focusing on long lasting marriages as positively framed, irrespective of whether the likelihood for the event to happen is high or low (which would impact the fact’s consequentiality).

In Hilbig’s (Hilbig, 2009, 2012a, b) experiments, participants read one version of the statement and were then asked to judge the truth of the stated information. Results indicate that

negatively framed statements are more likely to be evaluated as true than the content-wise identical, but positively framed statement (Hilbig, 2009, 2012b). This bias was found to be robust, with medium to large effect sizes ($ds = .60 - .82$; Hilbig, 2009).

But why is negatively framed information compared to positively framed information deemed more true? Negative information compared to positive information has a greater impact on human cognition, affect, and behavior in general (Baumeister, Bratslavsky, Finkenauer, & Vohs, 2001), which was coined the “Bad is stronger than good effect”. The authors argue, for example, that negative life events are far more influential than positive life events, as one negative major life event can have detrimental psychological effects (leading, e.g., to posttraumatic stress disorders), whereas multiple positive events cannot provide a sufficient buffer for individuals. This research is also consistent with more general evidence suggesting that losses compared to gains loom larger; meaning that framing an outcome in terms of gains or losses critically impacts individuals decisions and risk taking behavior (Kahneman & Tversky, 1984; Tversky & Kahneman, 1981).

Moreover, negative instances attract more attention (Pratto & John, 1991) and are perceived as more informative (Peeters & Czapinski, 1990), perhaps because negative instances are more rare and more threatening (Dijksterhuis & Aarts, 2003; Lewicka, Czapinski, & Peeters, 1992; Peeters & Czapinski, 1990). Interestingly, that negativity attracts more attention is also apparent in recent research on emotive intensifiers (e.g., such as “terribly” nice), suggesting that emotive intensifiers often relate to negative emotions (e.g., fright and horror; Jing-Schmidt, 2007) and therefore might result in closer attention to the information shared.

Another potential explanation for the “bad is stronger than good phenomenon” (Baumeister et al., 2001) stems from the observation that positive information is more similar than negative information (Alves, Koch, & Unkelbach, 2017; Koch, Alves, Krueger, & Unkelbach, 2016). Presumably this is because positive states are comparably non-extreme, whereas negative information is more distinct (Alves et al., 2017). To illustrate, for many

attribute dimensions, the extremes are negative, and the middle range is positive – individuals typically rate an average face (e.g., a morph of different faces) as more attractive than faces with extreme features (Langlois & Roggman, 1990). Against this background, individuals benefit more from paying special attention to negative information and from weighing it more strongly in their decision-making processes. As a result, negative (bad) information may be weighted more heavily than positive (good) information (Baumeister et al., 2001).

Using a multinomial processing tree model, Hilbig (2012b) argues that the negativity bias in judgments of truth is a response bias and not due to differences in knowledge. More specifically, he argues that the observed evidence is compatible with the idea that the accumulation of confirming evidence may proceed more quickly for the negative (bad) compared to the positive (good) information frame and that these differences in processing fluency may drive framing effects (p. 46, and see Hilbig, 2012a).

More recently, Koch and Peter (2017) investigated the context in framing and truth judgments. More specifically, their research focuses on the negativity bias in the context of political messages as well as the trustworthiness of politicians. The authors showed that positive framing can be associated with persuasive communication, and therefore reduce the perceived truth of the message as well as the trustworthiness of the source (Koch & Peter, 2017). All in all, different lines of research converge in suggesting that negative statements are more likely to attract attention, are more likely to be elaborated, and finally, are more likely believed as true.

In this context, we aimed at investigating the negativity bias in judgments of truth in an arena important for many individuals: health and well-being. In particular, we focused on health and well-being topics that were important to women in German speaking countries. To our surprise, the results of Study 1 were not consistent with earlier evidence on the negativity bias. We therefore proceeded by analyzing individuals' expectation regarding the specific contents (see Study 2) and found systematic patterns. Systematically adjusting the percentage rate in the

negative frame (Study 3) and in both frames (Study 4) then allowed investigating when a negativity bias in judgments of truth is likely to occur.

Study 1

Study 1 serves as a starting point for the current research line. The basic aim of Study 1 was to create a new set of items related to topics such as health and well-being that can be used to investigate the negativity bias. We hypothesized:

Hypothesis 1 (H1): Negatively compared to positively framed statements are more likely judged as true.

Given the bias's robustness in previous research across different statements, using new and previously untested statements did not occur to us as problematic. We constructed eight items that followed the structural logic introduced by Hilbig (2009). The items were related to German-speaking women, covering topics such as their status regarding health, well-being, and social security. An exemplary item is "70% of German-speaking women do not text while driving a car" (positive frame) or "30% of German-speaking women text while driving a car" (negative frame). All of the eight items were factually true.

Study 1 consists of two data sets that were collected within the frame of two months. The first set of data consists of 161 participants (Study 1a), the second set of data of 21 participants (Study 1b). As the two studies were highly similar in methodology, we combined the two datasets for analysis. Tables 2 and 4 (Appendix A) provide an overview of the findings separately for each dataset.

Method

Participants and Design. One hundred eighty-two students in a German-speaking country participated (61 male, 117 female, 4 no answer; $M_{age} = 23.71$, $SD_{age} = 5.62$) and judged the veracity of either eight positively or negatively framed items (between subjects factor: valence). Judgments of truth served as a dependent variable and were coded as 0 for *false* and 1 for *true*, and were averaged across the statements. In the Study 1b, we additionally and exploratorily asked participants to judge how threatening they perceived each statement's content to be on a seven-point Likert scale (1 = not threatening at all, 7 = very threatening). Participants received a snack and/or a small amount of monetary compensation (about 1 US\$).

Materials and Procedure. The eight items used in this study are presented in Table 1. The survey was conducted using tablet computers in the main university building. Research assistants approached passing students and asked them to participate in the study. Participants were asked to first read an exemplary statement, and to judge whether they believed this statement to be true or false (not analyzed). Participants were further told that not knowing the correct answer is perfectly normal, and were asked to make an educated guess, without using additional sources such as Google or Wikipedia. Participants then evaluated the truthfulness of eight positively or negatively framed items, and completed some demographic questions.

Results

When looking at descriptive results, across both datasets, we find that participants were more likely to judge the positive frame (compared to the negative frame) as true. Table 1 summarizes these results. Contrary to previous results in the literature, the positively framed items were more likely to be judged as true compared to the negatively framed items, $M = .56$, $SD = .20$; $M = .50$, $SD = .16$; respectively, $t(172.14) = 2.30$, $p = .023$, $d = 0.33$. The results therefore do not provide support for the occurrence of a negativity bias, but instead hint towards a positivity bias tendency.

< Insert Table 1 here >

Discussion

Study 1 failed to replicate the negativity bias in judgments of truth (Hilbig, 2009, 2012b) and did not provide support for H1. Instead we found a positivity bias. We took this unexpected finding as grounds to further explore the circumstances in which a negativity bias occurs, and started our search by comparing our study with existing ones. One possibility could be systematic item specificities that differentiate our items from the ones previously used. Yet no such systematic specificities were apparent to us. However, we realized that for those few items that showed a negativity bias in our sample, we were ourselves (positively) surprised about the stated truth. That is, some of the negative statements were less negative than we had expected—apparently, the world was in a better state than we had thought. To illustrate, learning that only 7% of women in German-speaking areas evaluate their health as bad or very bad appeared to us to as rather good news. Interestingly, when looking back at the items employed by Hilbig (2009), at least for some of them, the negatively framed truth surprised us in a positive sense, too. For instance, that only 20% of marriages result in divorce within the first 10 years in Germany surprised us in a positive sense, given that the average divorce rate in the authors' country of residence is slightly higher than 40% (see Federal Statistical Office, www.bfs.admin.ch). Consistent with findings in regards to the optimism bias (Sharot, Korn, & Dolan, 2011; Sharot, Riccardi, Raio, & Phelps, 2007), which holds that individuals are more likely to update their beliefs when learning that certain risks are lower (compared to higher) than expected, it is conceivable that individuals are biased to perceive good news as true, and therefore evaluate particularly those negatively framed statements as true that are more positive than expected.

Consistent with previous research showing that individuals' expectations impact judgments of truth more generally (e.g., Levine et al., 2000) and that expectations even

influence how language is perceived (Niedzielski, 1999), we hypothesize that individuals' expectations play a crucial role in the negativity bias, since expectations might critically influence whether a negative thing is negative, or in fact perceived as positive in regards to the subjectively construed consequentiality.

Study 2

To gain insight into individuals' expectations regarding the statements' content used in Study 1, we asked individuals to estimate the percentage of German-speaking women for whom a certain statement would be true.

Method

We used the same eight items as in the previous studies, and asked participants to estimate the percentage rates either for eight positively or negatively framed items (between subjects factor). Study 2 was conducted with tablet computers in the main university building. Research assistants approached passing students and asked them to participate in the short study for a chocolate bar. Eighty-one students participated (40 male, 41 female; $M_{age} = 25.44$, $SD_{age} = 6.06$) and completed the study.

Results

Individuals' percentage expectations are summarized in Table 2. In general, individuals' expectations differed from the factually true percentage rates. More particularly, estimates differed from the actual percentages by about 12 %-points for negatively framed items, and by about 14 %-points for positively framed items (absolute values).

< Insert Table 2 here >

Next, we investigated whether descriptively systematic patterns occurred on the item level. To this end we combined the results from Studies 1a and 2. Revisiting the data from Study 1a, a positivity bias occurred for the statements 1, 2, 3, and 4 (marked with an a in Table 2). For statements 5, 6, and 8 (marked by a b), a negativity bias was descriptively present. Coupled with the expectation ratings assessed in Study 2, an interesting pattern emerges. For items that showed a negativity bias, participants overestimated the percentage rates for the negatively framed items and underestimated the percentage rates for positively framed items. For items that showed a positivity bias, no consistent pattern emerges—some percentages were overestimated, some underestimated, and this was true for both the positive and the negative frame.

Discussion

In Study 2 we collected further data regarding the statements used in the previous study. Participants' mean expectations differed markedly from the true values. More importantly, however, a systematic pattern emerged as a function of over-/underestimation and truth judgments when comparing data across Studies 1a and 2. A negativity bias was more likely to be apparent in participants' truth judgments when the occurrence of the aspect was overestimated in the negatively framed version, and underestimated in the positively framed version. This pattern provides initial support for H2.

Study 3

The combination of the previous studies suggests that a negativity bias may be likely in cases where individuals overestimate the occurrence of a negatively framed fact, and underestimate the occurrence of a positively framed fact. This conclusion was reached in a data-driven fashion and resulted in Hypothesis 2.

Hypothesis 2 (H2): A negativity bias in judgments of truth is more likely to be present when the occurrence of the aspect is overestimated in the negatively framed version, and underestimated in the positively framed version.

To test H2 more rigorously, we now aimed to investigate in Study 3 whether we could *produce* a negativity bias in our set of eight statements regarding German-speaking women given the previous insights. Because the negativity bias highlights the role of the negative, we here focus on the negative frame. Given that participants judged statements as more likely to be true when they overestimated the frequency of occurrence in the negatively framed version, we adjusted the percentage rates for our statements in a systematic way as a function of statement valence. In particular, we changed the percentages so that for negatively framed items the new percentage rates were lower than expected (triggering overestimation), by subtracting an average of 13.5 points (the mean deviation between individuals' expectations and the actual percentages for the negatively framed statements 5 and 6, where a negativity bias occurred and strong deviations were apparent), from the previous item value. For positively framed items, we adjusted the percentages so that percentages across negative/positive framing add up to 100 (following Hilbig, 2009). If by this logic a percentage rate would have ended up being 0 or 100, we adjusted the percentage to 1 or 99 accordingly. Item 1 now read "10% (= expected value: 23.17% - constant: 13.5) of all German-speaking women die from cardiovascular diseases" in the negative frame and "90% (= 100 - adjusted percentage rate in negative frame: 10%) of all German-speaking women do not die from cardiovascular diseases" in the positive frame.

Method

Participants. Using G*Power (Faul, Erdfelder, Lang, & Buchner, 2007), we conducted a power analysis with an estimated medium effect size ($d = .5$), an alpha level of .05, and an aspired

power of 0.8. The analysis indicated a required sample size of 102 participants. The study was conducted in our research lab as part of a study package and was advertised in psychology lectures as well as on local online platforms. The study package took about 11 minutes to complete. One hundred fourteen individuals completed the questionnaire (24 males, 89 females, 1 no answer; $M_{age} = 21.99$ years, $SD_{age} = 6.06$). Participants received course credit as compensation.

Design. Participants read eight statements that differed in regards to valence (negative versus positive frame, between-subjects design). The dependent variable was the average perceived veracity over the eight items. Participants were randomly assigned to one of the two conditions.

Materials and Procedure. The Setup of Study 3 is equivalent to Studies 1a and 1b, except for the implementation of adjusted percentage values.

Results

Overall, participants judged 3.93 statements as true (about 49%; $SD = 1.49$), and individuals ranged from a minimum of 0 to a maximum of 7 statements judged as true. Looking at the statement-level, the average ratings of truthfulness varied from .38 to .58 for the individual items. As no statements seemed to be judged as true or false by all participants, we summarized the judgments and calculated a mean perceived veracity score over all eight statements.

To analyze if a negativity bias occurred, we calculated a t-test with valence as an independent variable and mean perceived veracity as a dependent variable. Our results yield a significant difference in mean perceived veracity depending on valence framing, $t(111.90) = -3.27$, $p = .001$, $d = .61$. Looking at mean perceived veracity scores, participants rated negatively framed statements as true on average in 55% of the cases ($SD = 18$), compared

to 43% of the cases for positively framed statements ($SD = .18$). Obviously, a negativity bias occurred – the negatively framed statements were more likely to be judged as true.

Discussion

Study 3 highlights the importance of individuals' expectations for the occurrence of a negativity bias. Compared to the previous studies, we adjusted the percentage rates so that the percentages of the negatively framed versions were lower than expected by a separate sample (assessed in Study 2). This simple change yielded a significant negativity bias where we had observed none or even a positivity bias in Studies 1a and 1b, and therefore provides support for H2.

In this study, we adjusted the percentages so that in the negatively framed version, the percentages were lower than expected. In the positively framed version, we calculated percentages as a function of the negative frames, so that positive and negative frames together add up to 100%. This procedure, however, does not allow for the systematic investigation of effects of over and underestimation for both negative and positively framed statements. We address this question in Study 4.

Study 4

In Study 4 we again focused on the effects of individuals' expectations on subsequent judgments of truth, but additionally focused on the impact of over- and underestimation within positive frames. Results from Study 3 point to the conclusion that percentages lower than expected in the negative frame may result in a negativity bias. Because we constructed the positive frames in Study 3 to be, content-wise, identical (percentage rates were mirrored to the adjusted rates in the negative frames), we were not able to systematically test the impact of expectations in the positive valence framing condition. Although previous research based on the propositions of the Expectancy Violation Theory (Burgoon & Jones, 1976) has not found evidence for valence asymmetries in the violation of expectations in impression formation (e.g.,

Brannon & Gawronski, 2018), it appeared commendable to systematically test both over- and underestimation, for both negative and positive frames in judgments of truth. As this objective comes with the methodological caveat that statements are not content-wise identical anymore across negative/positive framing, we compare effects of over- and underestimation within positive and negative frames.

Method

Participants. We conducted an a priori power analysis using G*Power (Faul et al., 2007) with an estimate of a medium effect size ($f = .25$), an alpha-level of .05, and an aspired power of .95, which resulted in a required sample size of 210 participants. Compared to Study 3 we aimed for a higher power, to account for an equal importance of both the alpha and the beta-error. We collected data from 234 participants (132 male, 102 female, $M_{age} = 38.63$, $SD_{age} = 12.40$) via the platform clickworker.de. Participants worked on the study for about 3 minutes and received 0.45 € (about 0.5 \$) as compensation.

Design. Participants judged the truthfulness of eight statements, which varied in regards to valence (a positive versus negative frame; between subjects factor) and direction of deviance (over- versus underestimation; between subjects factor). This results in a 2 (valence) * 2 (deviance in estimation) design, with average perceived veracity as the dependent variable.

Materials. We used the same eight statements as before, with a negative or positive frame. However, we adjusted the percentages so that in both valence conditions, the rates would be lower or higher than the previously assessed expectations (Study 2). Specifically, we subtracted or added one standard deviation to achieve percentages that were below or above the expected values (standard deviations were taken from Study 2). We did this adjustment separately for the negative and the positive frame to be able to test the impact of over- and underestimation

for both framing conditions. This procedure comes with the downside that positive versus negative frames are no longer content-wise equal, as the separately calculated percentage rates across the negative and the positive valence no longer add up to 100. Table 3 summarizes the set of new items.

Procedure. Study 4 follows the same procedure as Study 3, except for the adjusted percentages in the statements.

< Insert Table 3 here >

Results

Overall, participants judged 4.21 statements as true (about 53%; $SD = 1.67$), and individuals ranged from a minimum of 0 to a maximum of 8 statements judged as true. At the level of individual statements, the average ratings of truthfulness varied from .45 to .64. As no statement was judged as true or false by all participants, we summarized the judgments and calculated a mean perceived veracity score over all eight statements.

To investigate the interplay between valence and deviance in estimation, we calculated a 2 * 2 ANOVA with mean perceived veracity as the dependent variable. Interestingly, we find a main effect for deviance in estimation, $F(1, 230) = 5.16, p = .024, \eta_p^2 = .02$, indicating that statements were more likely to be believed as true in cases of overestimation compared to underestimation ($M = .56, SD = .20; M = .50, SD = .22$; respectively). This main effect was further qualified by an interaction between deviance in estimation and valence framing, $F(1, 230) = 6.05, p = .015, \eta_p^2 = .03$. Simple main effects indicate that for the negatively framed statements, overestimation compared to underestimation led to a higher likelihood of truth ratings, $M_{overestimation} = .57, SD = .21; M_{underestimation} = .45, SD = .20; F(1, 230) = 11.20, p = .001, \eta_p^2 = .05$. For positively framed statements no differences in truth ratings were apparent,

$M_{overestimation} = .54, SD = .18; M_{underestimation} = .55, SD = .23; F(1, 230) = 0.02, p = .895, \eta_p^2 = .00.$

The main effect for valence was not significant, $F < 1.5.$

Discussion

The results from Study 4 show that the deviance of the percentages in relation to expectations assessed in a separate sample (Study 2) play a crucial role when judging a statement's veracity. In Study 4, we find that overestimation generally appeared to increase the likelihood of truth ratings, yet this main effect is qualified by an interaction. Study 4, therefore, supports the hypothesis developed from Studies 2 and 3; for negatively framed items, percentages that are lower than expected result in higher levels of truth judgments. This difference did not occur for positively framed statements. As briefly discussed before, we used statements about rather important topics such as issues regarding health and well-being. We therefore speculate that in the negative frames, overestimation might simply feel better, as the world appears better than expected. Learning that 31% of women are dissatisfied with their looks while expecting the percentage rate to be much higher (54%) is likely good news. In contrast, with positive frames, overestimation does not feel better. Learning that 24% of women are satisfied with their looks, while expecting this number to be around 45%, is bad news and likely unsatisfactory. Interestingly, these findings support the conclusion made by Brannon and Gawronski (2018) that the negativity bias (in general, not specific to truth) might be a "conditional phenomenon", suggesting that "the impact of valenced information is not determined in an absolute sense but instead depends on people's broader sets of beliefs and expectancies" (p. 200; see also Skowronski & Carlston, 1989). This dovetails with our evidence that when information is framed negatively, individuals' expectations impact whether they believe the statement to be true or false.

General Discussion

In this manuscript we investigated the negativity bias in judgments of truth, which holds that negatively framed information (compared to content-wise identical, but positively framed information) is more likely to be deemed true. Four studies indicate that there might be more to tell about the negativity bias in judgments of truth than has been previously thought. In Study 1, we show that when using newly developed statements that follow the same logic inherent to those previously used (see Hilbig, 2009, 2012a, b), no support could be garnered for the occurrence of a negativity bias. If anything, the data points into the direction of a positivity bias. Upon closer inspection of our and previous items, it occurred to us that individuals' expectations may play a crucial role when judging the veracity of statistical statements. In Study 2 we therefore assessed individuals' expectations regarding the frequency of occurrence for all statements and observed an interesting pattern: When looking at statements that showed a negativity bias, participants had overestimated the frequency of events for the negatively framed statements. We think that this makes sense from a perspective guided by expectations: Learning that the actual percentages are lower than expected for something negative is ultimately good news (the world is better than expected) and therefore individuals presumably want to believe it to be true (Sharot et al., 2011, 2007). In turn, learning that the actual percentages are higher than expected for something negative is bad news (the world is worse than expected) and presumably individuals therefore would prefer this statement to be false.

Based on this data-driven speculation, we systematically tested the role of expectations by manipulating the percentages in our statements: In Study 3 we lowered the percentages in the negatively framed version to create the effect of learning that the world is better than expected. This resulted in a negativity bias where no one was observed in Studies 1a and 1b. In Study 4, we took it one step further and varied the percentages for both the negatively and the positively framed statements, to investigate over- and underestimation effects for both frames. Study 4 suggests that overestimation generally appeared to increase the likelihood of truth ratings. Importantly, this main effect was qualified by an interaction with valence, suggesting

that overestimation results in higher truth ratings only when the statements were framed negatively. Learning that the world is worse or better than expected appears to influence truth judgments more when looking at the downsides and negative states, and not so much at the positive states of our society.

Importantly, throughout the studies we adjusted the percentages according to sample mean estimations of frequency assessed in Study 2. Therefore, we assume that over- and underestimation was present on the group level, however, this might have not been the case for every single individual. As a result, effects of expectations might be even stronger when considering individuals' idiosyncratic expectations. This procedure has been used in other areas of research, which also hinted that the direction of a deviance between observation and expectation plays a crucial role (and more so when the deviance feels positive compared to negative). For example, Sharot, Korn, and Dolan (2011) presented participants with 80 descriptions of adverse life events and asked how likely this event was to occur to them. After providing an estimate, individuals were presented with the average probability and then, in a second session, individuals were again asked how likely it was to them that such an event could occur. Although the researchers found that participants remembered descriptions for which they over- and underestimated the likelihood equally well, they updated their beliefs in an asymmetric way. Participants updated their beliefs more in response to information for which they overestimated the occurrence (percentages were better than expected) compared to information for which they underestimated the occurrence (percentages were worse than expected). This tendency was coined "optimism bias" (Sharot et al., 2007). Using such a two-sessions-paradigm, future research could investigate whether effects of over- versus underestimation might be even stronger when taking idiosyncratic expectations into account.

The present research also allows for first assumptions regarding the processes that drive the effects of over- and underestimation of negatively and positively framed statements. As noted above, learning that the actual percentages are lower than expected for something

negative could be perceived as good and therefore individuals presumably want to believe it to be true. In turn, learning that the actual percentages are higher than expected for something negative is bad news, which may trigger a preference for this statement to be false. Believing in good news and discarding bad news might be a helpful tendency, as previous research has highlighted that these optimistic expectations can promote better health and well-being (Taylor & Brown, 1988). Similar logic could explain why, in Studies 1a and 1b, we found a positivity bias compared to a negativity bias. While negatively framed information might still attract more attention (in line with the “bad is stronger than good” phenomenon, Baumeister et al., 2001), individuals could still be motivated to see the world as a good place. When being confronted with the negatively framed version of all our statements, participants might have been motivated to keep an optimistic view and discard statements when they were considered bad news. This nicely aligns with the research on the optimism bias, where individuals update their beliefs when learning about potential health risks – but they do so more strongly when the new information paints a more desirable picture of the world they are living in. Looking back at the estimated percentage rates from Study 2, we find that in the original negatively framed statements participants only overestimated the occurrence of three out of eight, which indicates that most of them were still bad news. Changing this situation in Studies 3 and 4 then led to a shift and also to a higher likelihood that the negative compared to the positive framing was deemed true. Nevertheless, further research is required to deep dive into the specific mechanisms and investigate this assumption, as the present studies set out to explain when a negativity bias is likely, but not why a positivity bias occurs.

Future research may also aim to replicate and broaden our findings with different statements focusing on a broader variety of topics. All statements employed in this contribution focused on health and well-being. However, other work investigating framing effects in the health domain show inconclusive findings, sometimes suggesting no differences, for example for promoting vaccination (O’Keefe & Nan, 2012), an advantage for gain-frames appeals when

advocating dental hygiene (O’Keefe & Jensen, 2007), yet an advantage for loss-frames appeals when advocating breast detection behaviors (O’Keefe & Jensen, 2009). We focused on health and well-being in our studies, because there is high consensus about what is good (good health and well-being) versus bad (bad health and well-being). This divide in good/bad is far more debatable for many other topics, and might also vary across individuals; another aspect that future research could investigate by focusing more on individual differences in expectations and judgments.

Next to different topics, it could also be interesting to look at the percentage rates and on the numbers’ actual impact on participants’ reactions. Taking a different approach than we did in our manuscript, one could keep the percentage rates between negative and positive frames constant and in the most extreme case at 50%. Ideally, researchers would find statements for which such a distribution is plausible and in alignment with individuals’ expectations. Then framing effects could be investigated while slowly adjusting the percentage rates, so as to see whether there are any systematic interactions between framing and numbers presented.

Furthermore, different types of statements could be investigated, as framing effects could differ depending on the type of statement. For instance, in contrast to statistical statements, advice appears to be evaluated as qualitatively superior when framed positively (in terms of gains) compared to negatively (in terms of losses; Jang & Feng, 2018). In addition, the intensity of language used (such as in utterances) might also play a crucial role in language perception and might differ for positively and negatively framed information (Liebrecht, Hustinx, & van Mulken, 2019).

Last but not least, research could also investigate different aspects of negativity. In this research, we focus on the framing of the target concept. However, one could investigate how negations, whether they are implicit or explicit, affect response biases in the same or a different setup. Holleman, Kamoen, Krouwel, van de Pol, and de Vreese (2016), for instance, have shown negation matters in online voting advice applications.

This manuscript highlights the general importance of expectations when making judgments of truth. This relation is especially important when being confronted with communication in a world of fake news and post-truth politics. In the worst case, a false statement is communicated in a way that mocks credibility by systematically playing off individuals' expectations. In a protective fashion, the knowledge derived from the results of the studies in this manuscript should be used to carefully investigate the framing of a statement. Have the percentages been chosen to exceed or fall below expectations? Has the framing been manipulated to specifically highlight the negative or positive aspects of a cause? How does reading such a statement make me feel? Questions like these might be asked to activate a mindful handling of ambivalent information: Somebody might try to make the statement feel good, driving a “less than expected and oh, so true” reaction in unjustified cases.

Looking back at our initial example of political information communicated via Twitter, the present results suggest that framing and expectations need to work in tandem to increase believability. That is, framing a statement as 51% approval rate versus a 49% disapproval may not be enough; it is important that when the negatively framed message is presented, the information is less negative than expected.

Acknowledgements

The authors would like to thank Howard Giles and the two anonymous reviewers for their valuable feedback on earlier drafts of this article. We would also like to thank Caroline Tremble for proofreading our manuscript and Jonas Mumenthaler, Claudia Puttinger, and Daniela Sutter for their help with the programming of the studies.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

Funding for the studies within this manuscript has been provided by the Center for Social Psychology, University of Basel, Switzerland.

References

- Alves, H., Koch, A., & Unkelbach, C. (2017). Why good is more alike than bad: Processing implications. *Trends in Cognitive Sciences, 21*, 69–79. doi:10.1016/j.tics.2016.12.006
- Baumeister, R. F., Bratslavsky, E., Finkenauer, C., & Vohs, K. D. (2001). Bad is stronger than good. *Review of General Psychology, 5*, 323–370. doi:10.1037//1089-2680.5.4.323
- Brannon, S. M., & Gawronski, B. (2018). In search of a negativity bias in expectancy-violation. *Social Cognition, 36*, 199–220. doi:10.1521/soco.2018.36.2.199
- Burgoon, J. K., Blair, J. P., & Strom, R. E. (2008). Cognitive biases and nonverbal cue availability in detecting deception. *Human Communication Research, 34*, 572–599. doi:10.1111/j.1468-2958.2008.00333.x
- Burgoon, J. K., & Jones, S. B. (1976). Toward a theory of personal space expectations and their violations. *Human Communication Research, 2*, 131–146. doi:10.1111/j.1468-2958.1976.tb00706.x
- Clementson, D. E. (2018). Truth bias and partisan bias in political deception detection. *Journal of Language and Social Psychology, 37*, 407–430. doi:10.1177/0261927X17744004
- Dechêne, A., Stahl, C., Hansen, J., & Wanke, M. (2010). The truth about the truth: A meta-analytic review of the truth effect. *Personality and Social Psychology Review, 14*, 238–257. doi:10.1177/1088868309352251
- Dijksterhuis, A., & Aarts, H. (2003). On wildebeests and humans: The preferential detection of negative stimuli. *Psychological Science, 14*, 14–18. doi:10.1111/1467-9280.t01-1-01412
- Dunwoody, P. T. (2009). Theories of truth as assessment criteria in judgment and decision making. *Judgement and Decision Making, 4*, 116–125.
- Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior*

Negativity Bias and Judgments of Truth: *Jaffé and Greifeneder*

Research Methods, 39, 175–191. doi:10.3758/BF03193146

Hilbig, B. E. (2009). Sad, thus true: Negativity bias in judgments of truth. *Journal of Experimental Social Psychology*, 45, 983–986. doi:10.1016/j.jesp.2009.04.012

Hilbig, B. E. (2012a). Good things don't come easy (to mind): Explaining framing effects in judgments of truth. *Experimental Psychology*, 59, 38–46. doi:10.1027/1618-3169/a000124

Hilbig, B. E. (2012b). How framing statistical statements affects subjective veracity: Validation and application of a multinomial model for judgments of truth. *Cognition*, 125, 37–48. doi:10.1016/j.cognition.2012.06.009

Holleman, B., Kamoen, N., Krouwel, A., Van De Pol, J., & De Vreese, C. (2016). Positive vs. Negative: The impact of question polarity in voting advice applications. *PLoS ONE*, 11, 1–17. doi:10.1371/journal.pone.0164184

Jang, J., & Feng, B. (2018). The effects of gain- and loss-framed advice messages on recipients' responses to advice. *Journal of Language and Social Psychology*, 37, 181–202. doi:10.1177/0261927X17706961

Jing-Schmidt, Z. (2007). Negativity bias in language: A cognitive-affective model of emotive intensifiers. *Cognitive Linguistics*, 18, 417–443. doi:10.1515/COG.2007.023

Kahneman, D., & Tversky, A. (1984). Choices, values, and frames. *American Psychologist*, 39(4), 341–350. doi:10.1037/0003-066X.39.4.341

Koch, A., Alves, H., Krueger, T., & Unkelbach, C. (2016). A general valence asymmetry in similarity: Good is more alike than bad. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 42, 1171–1192. doi:10.1037/xlm0000243

Koch, T., & Peter, C. (2017). Effects of Equivalence Framing on the Perceived Truth of Political Messages and the Trustworthiness of Politicians. *Public Opinion Quarterly*, 81, 847–865. doi:10.1093/poq/nfx019

Langlois, J. H., & Roggman, L. A. (1990). Attractive faces are only average. *Psychological*

Science, 1, 115–121. Retrieved from <http://www.jstor.org/stable/40062595>

Levine, T. R., Anders, L. N., Banas, J., Baum, K. L., Endo, K., Hu, A. D. S., & Wong, N. C.

H. (2000). Norms, expectations, and deception: A norm violation model of veracity judgments. *Communication Monographs*, 67, 123–137.

doi:10.1080/03637750009376500

Lewicka, M., Czapinski, J., & Peeters, G. (1992). Positive-negative asymmetry or “When the heart needs a reason.” *European Journal of Social Psychology*, 22, 425–434.

doi:10.1002/ejsp.2420220502

Liebrecht, C., Hustinx, L., & van Mulken, M. (2019). The relative power of negativity: The influence of language intensity on perceived strength. *Journal of Language and Social Psychology*, 38, 170–193. doi:10.1177/0261927X18808562

Niedzielski, N. (1999). The effect of social information on the perception of sociolinguistic variables. *Journal of Language and Social Psychology*, 18, 62–85.

doi:10.1177/0261927X99018001005

O’Keefe, D. J., & Jensen, J. D. (2007). The relative persuasiveness of gain-framed and loss-framed messages for encouraging disease detection behaviors: A meta-analytic review. *Journal of Health Communication*, 12, 623–644. doi:10.1080/10810730701615198

O’Keefe, D. J., & Jensen, J. D. (2009). The relative persuasiveness of gain-framed and loss-framed messages for encouraging disease detection behaviors: A meta-analytic review. *Journal of Communication*, 59, 296–316. doi:10.1111/j.1460-2466.2009.01417.x

O’Keefe, D. J., & Nan, X. (2012). The Relative Persuasiveness of Gain- and Loss-Framed Messages for Promoting Vaccination: A Meta-Analytic Review. *Health Communication*, 27, 776–783. doi:10.1080/10410236.2011.640974

Oxford Dictionary. (2016). *Oxford Dictionary Word of the Year 2016*. Retrieved July 15, 2019, from <https://languages.oup.com/word-of-the-year/word-of-the-year-2016>

Peeters, G., & Czapinski, J. (1990). Positive-negative asymmetry in evaluations: The

- distinction between affective and informational negativity effects. *European Review of Social Psychology*, *1*, 33–60. doi:10.1080/14792779108401856
- Pratto, F., & John, O. P. (1991). Automatic vigilance: The attention-grabbing power of negative social information. *Journal of Personality and Social Psychology*, *61*, 380–391. doi:10.1037/0022-3514.61.3.380
- RealDonaldTrump. (2018). Rasmussen just came out at 51% Approval despite the Fake News Media. [...] [TwitterPost]. Retrieved May 17, 2018, from <https://twitter.com/realDonaldTrump>
- Reber, R., & Schwarz, N. (1999). Effects of perceptual fluency on judgments of truth. *Consciousness and Cognition*, *8*, 338–342. doi:ccog.1999.0386
- Reder, L. M. (1982). Plausibility judgments versus fact retrieval: Alternative strategies for sentence verification. *Psychological Review*, *89*, 250–280. doi:10.1037//0033-295X.89.3.250
- Sharot, T., Korn, C. W., & Dolan, R. J. (2011). How unrealistic optimism is maintained in the face of reality. *Nature Neuroscience*, *14*, 1475–1479. doi:10.1038/nn.2949
- Sharot, T., Riccardi, A. M., Raio, C. M., & Phelps, E. A. (2007). Neural mechanisms mediating optimism bias. *Nature*, *450*, 102–105. doi:10.1038/nature06280
- Skowronski, J. J., & Carlston, D. E. (1989). Negativity and extremity biases in impression formation: A review of explanations. *Psychological Bulletin*, *105*, 131–142. doi:10.1037/0033-2909.105.1.131
- Taylor, S. E., & Brown, J. D. (1988). Illusion and well-being: A social psychological perspective on mental health. *Psychological Bulletin*, *103*, 193–210. doi:10.1037/0033-2909.103.2.193
- Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science*, *211*, 453–458.

Table 1. Overview of Statements, Mean Perceived Veracity from Study 1

		% of “true” judgments		% of “true” judgments
	Negative frame		Positive frame	
1	40% of all German-speaking women die from cardiovascular diseases	27	60% of all German-speaking women do not die from cardiovascular diseases	68
2	61% of German-speaking women are dissatisfied with their looks	62	39% of German-speaking women are satisfied with their looks	67
3	30% of German-speaking women text while driving a car	48	70% of German-speaking women do not text while driving a car	62
4	35% of 18-25 year old women in German-speaking regions engage in binge-drinking	19	65% of 18-25 year old women in German-speaking regions only drink moderately	62
5	7% of women in German-speaking areas evaluate their health as “bad” or “very bad”	57	93% of women in German-speaking areas evaluate their health as “good” or “very good”	35
6	5% of German-speaking women currently suffer from a Burn-Out	62	95% of German-speaking women currently do not suffer from a Burn-Out	56
7	7% of working women in German-speaking areas take antidepressants	51	93% of working women in German-speaking areas do not take antidepressants	56
8	26% of single mothers in German-speaking areas depend on social welfare benefits	74	74% of single mothers in German-speaking areas do not need social welfare benefits	42

Table 2. Overview of Statements, Mean Perceived Veracity, and Estimated Applicability to German-Speaking Women from Study 1a and Study 2

	% of “true” judgments [Study 1a]	Estimated % rate [Study 2]	Difference estimation and fact		% of “true” judgments [Study 1a]	Estimated % rate [Study 2]	Difference estimation and fact	
	Negative frame				Positive frame			
1	40% of all German-speaking women die from cardiovascular diseases ^a	29	23.17	- 16.83	60% of all German-speaking women do not die from cardiovascular diseases ^a	68	60.38	0.38
2	61% of German-speaking women are dissatisfied with their looks ^a	59	54.05	- 6.95	39% of German-speaking women are satisfied with their looks ^a	65	44.98	5.98
3	30% of German-speaking women text while driving a car ^a	49	35.54	5.54	70% of German-speaking women do not text while driving a car ^a	62	62.03	- 7.97
4	35% of 18-25 year old women in German-speaking regions engage in binge-drinking ^a	19	16.15	-18.85	65% of 18-25 year old women in German-speaking regions only drink moderately ^a	59	57.05	-7.95
5	7% of women in German-speaking areas evaluate their health as “bad” or “very bad” ^b	55	19.95	12.95	93% of women in German-speaking areas evaluate their health as “good” or “very good” ^b	33	66.75	- 26.25
6	5% of German-speaking women currently suffer from a Burn-Out ^b	64	19.20	14.20	95% of German-speaking women currently do not suffer from a Burn-Out ^b	54	72.90	- 22.10
7	7% of working women in German-speaking areas take antidepressants	53	23.17	16.17	93% of working women in German-speaking areas do not take antidepressants	53	75.80	- 17.20
8	26% of single mothers in German-speaking areas depend on social welfare benefits ^b	75	30.73	4.73	74% of single mothers in German-speaking areas do not need social welfare benefits ^b	43	46.62	- 27.38

^a occurrence of a positivity bias; ^b occurrence of a negativity bias

Table 3. Overview of Statements, Expectation Ratings from Study 2 and Adjustments for Study 4

	Estimated % rate [Study 2]	Adj. % rates below expectations	Adj. % rates above expectations		Estimated % rate [Study 2]	Adj. % rates below expectations	Adj. % rates above expectations
Negative frame	<i>M (SD)</i>			Positive frame	<i>M (SD)</i>		
1 XX% of all German-speaking women die from cardiovascular diseases	23.17 (16.58)	7	40	XX% of all German-speaking women do not die from cardiovascular diseases	60.38 (20.00)	40	80
2 XX% of German-speaking women are dissatisfied with their looks	54.05 (23.08)	31	77	XX% of German-speaking women are satisfied with their looks	44.98 (21.40)	24	66
3 XX% of German-speaking women text while driving a car	35.54 (25.31)	10	61	XX% of German-speaking women do not text while driving a car	62.03 (22.34)	40	84
4 XX% of 18-25 year old women in German-speaking regions engage in binge-drinking	16.15 (15.22)	1	31	XX% of 18-25 year old women in German-speaking regions only drink moderately	57.05 (24.68)	32	82
5 XX% of women in German-speaking areas evaluate their health as “bad” or “very bad”	19.95 (13.50)	6	33	XX% of women in German-speaking areas evaluate their health as “good” or “very good”	66.75 (16.71)	50	83
6 XX% of German-speaking women currently suffer from a Burn-Out	19.20 (12.63)	7	32	XX% of German-speaking women currently do not suffer from a Burn-Out	72.90 (21.33)	52	94
7 XX% of working women in German-speaking areas take antidepressants	23.17 (13.65)	10	37	XX% of working women in German-speaking areas do not take antidepressants	75.80 (15.04)	61	91
8 XX% of single mothers in German-speaking areas depend on social welfare benefits	30.73 (19.61)	11	50	XX% of single mothers in German-speaking areas do not need social welfare benefits	46.62 (24.97)	22	72

Appendix A

Table 4. Overview of Statements, Mean Perceived Veracity from Study 1b

		% of “true” judgments			% of “true” judgments
Negative frame			Positive frame		
1	40% of all German-speaking women die from cardiovascular diseases	18	60% of all German-speaking women do not die from cardiovascular diseases		70
2	61% of German-speaking women are dissatisfied with their looks	82	39% of German-speaking women are satisfied with their looks		80
3	30% of German-speaking women text while driving a car	46	70% of German-speaking women do not text while driving a car		60
4	35% of 18-25 year old women in German-speaking regions engage in binge-drinking	18	65% of 18-25 year old women in German-speaking regions only drink moderately		80
5	7% of women in German-speaking areas evaluate their health as “bad” or “very bad”	73	93% of women in German-speaking areas evaluate their health as “good” or “very good”		50
6	5% of German-speaking women currently suffer from a Burn-Out	46	95% of German-speaking women currently do not suffer from a Burn-Out		70
7	7% of working women in German-speaking areas take antidepressants	36	93% of working women in German-speaking areas do not take antidepressants		80
8	26% of single mothers in German-speaking areas depend on social welfare benefits	64	74% of single mothers in German-speaking areas do not need social welfare benefits		30

Author Biographies

Mariela E. Jaffé is a Postdoctoral Researcher in Social Psychology at the University of Basel, Switzerland. Her research interests include decision making aids, diversity preferences, and judgments of truth.

Rainer Greifeneder is Professor of Social Psychology at the University of Basel, Switzerland. His research interests are centered around the impact feelings have on judgment and decisions, and on individuals' experience and perception of social ostracism.