A hotel that is not bad isn't good. The effects of valence framing and expectation in online reviews on text, reviewer and product appreciation

Naomi Kamoen [a,b,1,*], Maria B.J. Mos [a,1,*], Willem F.S. Dekker (Robbin) [a]

[a] Tilburg University, The Netherlands
[b] Utrecht University, The Netherlands

Abstract

In online hotel reviews, reviewers use both direct and indirect positive and negative evaluations (e.g. 'good', 'not bad', 'bad', 'not good'). In four studies, we examined the effects of these wording alternatives. In Study 1, participants rated hotel reviews that were manipulated with respect to the wording. In positive reviews, direct evaluations ('good') received higher evaluations than indirect wordings ('not bad'). In negative reviews, however, no such difference was observed between direct and indirect expressions ('not good'/'bad'). These results apply to evaluations of the hotel, text and reviewer alike. Study 2 showed that this pattern of results generalizes to restaurant reviews. To investigate an underlying cause for the effects in Study 1 and 2, we manipulated participants' a priori expectation of the attitude object (hotel) in two subsequent studies. The lack of an interaction effect between the wording and expectation manipulations shows that the pattern of results may be attributed to Verbal Politeness: wordings like 'not bad' convey a weakened meaning as compared to 'good', whereas the use of 'not good' instead of 'bad' is interpreted as expressing the same evaluation, albeit in a more polite way.
© 2014 Elsevier B.V. All rights reserved.

Keywords: Online reviews; Framing; Expectation; Verbal politeness; Argumentative orientation

1. Introduction

Before booking a hotel, more and more people consult online reviews (e.g. Gretzel and Yoo, 2008; Ye et al., 2009). Via these sources of Electronic Word of Mouth (Litvin et al., 2008), they check if earlier costumers found the service 'good' and the rooms 'comfortable'. If this is the case, they are likely to form a positive attitude and book a room. However, if earlier customers described the service as 'bad' and the rooms as 'uncomfortable', they will continue their search for better options (Ye et al., 2009; Vermeulen and Seegers, 2009).

But what happens when the service is described as 'not good' or when the rooms are deemed 'not uncomfortable'? Do these indirect descriptions evoke other impressions than their direct counterparts? And does the wording affect the review reader's judgment of the product that is reviewed as well as the text and its writer in equal ways? These questions have been investigated in the experiments presented in this article. Therefore, the current research provides insight into the effects of direct and indirect positive and negative language in a realistic language use setting in which evaluative descriptions are in abundance.

---

* Corresponding authors at: Tilburg University, Warandelaan 2, 5037 AB Tilburg, The Netherlands. Tel.: +31 013 466 9111.
   E-mail addresses: n.kamoen@uvt.nl (N. Kamoen), maria.mos@uvt.nl (M.B.J. Mos), dekkerwfs@gmail.com (W.F.S. Dekker (Robbin)).
   [1] These authors contributed equally to this study.

1.1. Framing

Often, a situation can be described in both positive terms (e.g. 'this medical treatment has a 50% success rate') and negative terms (e.g. 'this medical treatment has a 50% failure rate'). These alternative wordings are logically equivalent: they describe the same reality. Yet, the positive or negative frame in which information is presented influences our response with regards to the evaluation we make or decision we take. For example, subjects rate the medical treatment as more effective and they are more likely to recommend the treatment to others when it is described as having a '50% success rate' rather than a '50% failure rate' (Levin et al., 1988).

The medical treatment example above concerns a specific type of framing that Levin et al. (1998) call attribute framing: the framing of one object or event characteristic (i.e. the treatment) in evaluatively positive (i.e. success) or negative (i.e. failure) terms.[2] Attribute framing has typically been investigated in studies that look at effects in single occurrences, that is, one frame manipulation per stimulus. In addition, the framing manipulation often concerns contrary antonyms to which proportions can be assigned, such as percentages of success (Levin et al., 1988), or mutually exclusive antonyms, comparing such things as 'survival' versus 'mortality' rates (Marteau, 1989). The main outcome of these studies is a valence-consistent shift (Levin et al., 1998): the selection of a positive frame (e.g. a description of a percentage of 'success' or 'survival' rates) leads to more favorable evaluations than a negative frame (e.g. a description of a percentage of 'failure' or 'mortality' rates).

1.2. Why people choose one frame over the other frame

After the initial observation that frame choice influences interpretation, the question arises when and why people select a particular frame in speech or text production. In various experiments (e.g. McKenzie and Nelson, 2003; Sher and McKenzie, 2009), the choice of a particular frame has been shown to reveal speakers' reference points. The reference point, or baseline, as termed by Langacker (2008), is the starting point that serves as the background to the situation described. In positive/negative pairs, usually one of the two poles serves as the relatively neutral reference point and is the most commonly used frame in speech production (Tribushinina, 2008). For example, when one wants to ask about someone's achievements in the last five tennis matches, the more neutral question is 'How many tennis matches did you win?', rather than 'How many tennis matches did you lose?' (see also Horn, 1989). Hence, the 'winning' frame is in this case the more neutral frame, which is used by default. However, there may be situations in which another reference point becomes salient in the discourse. Experiments demonstrating that reference points are important in determining frame production choices, usually focus on such situations.

For example, in a study by McKenzie and Nelson (2003), speakers had to describe a 4-ounce cup filled to the 2-ounce line either when the cup was previously empty (condition 1), or when the cup had been full (condition 2). Across conditions, the 'half full' frame was chosen more often than the 'half empty' frame. This seems to suggest that 'half full' is the default frame to use. In addition to the general preference for 'half full', the cup was more often described as 'half full' when it was previously empty, but as 'half empty' when it had previously been full. Hence, the frame chosen depends on which proportion has increased/decreased. This indicates that speakers tailor their frames to a salient reference point in a discourse context, e.g. how the proportion has changed as compared to a previous situation (see also Sher and McKenzie, 2009).

Similar results have been found in a series of experiments by Holleman and Pander Maat (2009). In one of their studies, participants had to describe the results of a tennis player who had a good (condition 1) versus a bad (condition 2) year. When speakers knew the tennis player experienced a 'good year', they talked about the number of matches won. However, when speakers were informed that the tennis player had a 'bad year', they more often mentioned the number of matches lost. This shows that reference points may also include speakers' opinions: the direction of the frame corresponds with the object evaluation. This principle is called Argumentative Orientation (Holleman and Pander Maat, 2009).

1.3. Why responses are influenced by valence frames

Sher and McKenzie (2006) tie the observations for frame production and frame interpretation together. They argue that the effects found in frame interpretation can be explained from listeners inferring the speaker's reference point from their description. This is what they call information leakage: while ostensibly equivalent descriptions in frames can be logically equivalent, they need not be informationally equivalent, because the choice of frame 'leaks' information about the

---

[2] Levin et al. (1998) distinguish two other types of framing in addition to attribute framing: risky choice framing and goal framing. In risky choice framing "the outcomes of a potential choice involving options differing in level of risk are described in different ways" (p. 150). This is the type of framing investigated in the famous framing studies by Kahneman and Tversky (e.g. 1979, 1984). In goal framing the goal of an action or behavior is framed in different ways (e.g. Banks et al., 1995). We refer to Levin et al. (1998) for a further definition and examples of these different types of framing.

reference point of the speaker (e.g. an increasing proportion). The notion of information equivalence builds on Gricean maxims (Grice, 1975), in that listeners assume that what is expressed is relevant in its content. If a speaker chooses to frame a situation in terms of losing (e.g. 'Tim lost 40% of the tennis matches last week'), rather than winning (e.g. 'Tim won 60% of the tennis matches last week'), that must be relevant. Mentioning the number of losses is irrelevant at the level of logical equivalence – there are no ties in tennis matches: if you know the percentage of games won, you also know the percentage of games lost. Therefore the 'losing' frame must be indicative of something else, a reference point, such as an increasing proportion as compared to a previous situation (e.g. the week before Tim lost no matches at all).

Holleman and Pander Maat (2009) argue that the information leakage explanation ''is not really communicative'' as it ''suggests an automatic process, beyond the control of the language user'' (p. 2206). The concept of leakage seems to imply that the information that is provided through the use of one frame rather than another, is both involuntary and arbitrary. Holleman and Pander Maat therefore add to Sher and McKenzie's information leakage explanation that reference points are often chosen voluntarily, because they reflect speaker opinions, and thereby they are not arbitrary either. In other words, speakers will choose to produce frames that match their opinions and listeners will interpret a chosen frame as indicative of an evaluative stance taken by the speaker. Hence, the principle of Argumentative Orientation (AO) goes for both frame production and frame interpretation.

1.4. Attribute framing in online reviews

A corpus study shows that in the genre of online reviews, evaluative descriptions are in abundance and that review writers vary in the way they formulate these evaluations (Beks, 2012). For example, hotels are sometimes described as 'cheap' and other times as 'not expensive'. Negative evaluations may also be framed in different ways, as personnel may be called 'unfriendly' or 'not friendly'.

While these wording choices in some respect resemble the linguistic frames that have been studied before in framing research, there are at least two differences. First, in previous framing studies, the focus is often on contraries that carry a calibratable meaning, and therefore different frames can be constructed assigning proportions (e.g. '70% success', '30% failure'). The contrasts used in online reviews are contraries that cannot be calibrated in terms of percentages; e.g., a hotel cannot be '70% cheap'. Therefore, framing differences translate into the usage of direct and indirect wordings ('bad'/'not good'). A second observation is that in the context of online reviews, framing is important for both sides of the scale: both positive and negative evaluations can be framed using direct ('good', 'bad') and indirect ('not bad', 'not good') terminology. Therefore, an investigation of framing effects in a 2 (evaluative content: positive/negative) × 2 (form: direct/indirect) design is relevant; this is not the most common set-up in framing research.

Reasoning from the principle of AO, which can be applied to the argumentative context of an online review par excellence, the prediction is that indirect forms contain a weakened meaning as compared to their direct counterparts: if the speaker had had a truly positive or negative opinion, he would have used direct terminology. This prediction of framing effects based on AO is depicted in Fig. 1. We call this hypothesis derived from AO the *symmetrical mitigation hypothesis* (see Fig. 1).

1.5. The principle of verbal politeness

Outside the field of framing studies, extensive research has been done on the effects of direct and indirect positive and negative language on language interpretation. From various theoretical frameworks (see for example Horn, 1989), predictions about the effects of direct and indirect positive and negative language can be derived that are different from the *symmetrical mitigation hypothesis*.

One such hypothesis can be derived from the principle of Verbal Politeness (VP). This principle builds on the Positivity Bias, the tendency to be positive towards actions, objects, humans etcetera (Brown and Levinson, 1987; also see Colston, 1999). From this bias, politeness norms arise that prescribe people to speak in a socially acceptable positive manner. If speakers want to convey a negative evaluation, they can either use a negative word (e.g. 'bad') or negate the opposing positive term (e.g. 'not good'). Many negative terms have a negative connotation, leading people to use the

Good				Bad
	Not Bad		Not good	

Fig. 1. Relative distance in perceived evaluation of direct and indirect positive and negative terms according to the symmetrical mitigation hypothesis (cf. Holleman and Pander Maat, 2009).

Good						Bad
	Not Bad					Not Good

Fig. 2. Relative distance in perceived evaluation of direct and indirect positive and negative terms according to the positive evaluation mitigation hypothesis (cf. Brown and Levinson, 1987).

negated positive term instead of the direct negative term, so as not to be viewed as impolite. However, if the use of 'not good' is caused by Verbal Politeness, it may be interpreted as being 'just as negative' as its direct counterpart 'bad'. Speakers' selection of 'not bad' over 'good', on the other hand, cannot be for reasons of verbal politeness. Therefore, 'not bad' must indicate a different, slightly less positive evaluation than 'good'. This prediction based – solely – on the principle of Verbal Politeness has been depicted in Fig. 2. We call this prediction the *positive evaluation mitigation hypothesis*.

Two of the studies (experiment 1 and 3) reported in Colston (1999) may be considered empirical evidence for the *positive evaluation mitigation hypothesis*.[3] In these studies, Colston presented participants with a scenario in which someone provided an evaluation of a given situation. In this evaluation either direct positive ('good'), indirect positive ('not bad'), direct negative ('bad'), or indirect negative ('not good') descriptions were used. The participants' task was to indicate their opinion about the speaker's intended meaning. Results of Colston's study 1 and 3 showed that positive evaluations ('good', 'not bad') were seen as more positive than negative evaluations ('bad', 'not good') but, more importantly, there was no perceived difference between direct and indirect negative comments ('bad', 'not good'), whereas an indirect positive comment ('not bad') was seen as less positive than a direct positive comment ('good').

1.6. Goals of the current research

From AO and VP different hypotheses can be derived about how respondents interpret an object when it is described in direct or indirect positive or negative language. A first aim of the current research is to test these hypotheses in the naturalistic and argumentative context of an online hotel review.

A second aim of the current research is to explore whether or not direct and indirect positive and negative language use affects not only the evaluation of the object that is described in the review, but also the evaluation of the reviewer and the review itself. Some first thoughts on how these evaluations may be affected, can be derived from framing literature. This literature suggests that using indirect terminology is something that speakers/writers often do in natural language use. So, there is no reason to expect that framing affects the evaluation of the writer and the evaluation of the text in online hotel reviews.

2. Study 1

2.1. Method

2.1.1. Aim

In order to empirically contrast the *symmetrical mitigation hypothesis* and the *positive evaluation mitigation hypothesis*, we set up an experimental study in which participants evaluated online reviews in terms of hotel, text and reviewer appreciation.

2.1.2. Design

Participants were assigned to a 2 (form: direct versus indirect) × 2 (content: positive versus negative) design. Hotel reviews were written about four fictional hotels. For each hotel, four versions of a review were constructed (see Table 1). Participants saw one review of each hotel and one exemplar of each experimental condition. The combination of hotel review and condition was randomized.

2.1.3. Material

In order to create realistic reviews, a corpus study identified common topics and wordings in online hotel reviews. Based on the corpus study, each review was built up from the same elements: a neutral descriptive first sentence, four evaluative sentences with one manipulation each, two sentences with comments on other aspects of the holiday, and a

---

[3] Colston (1999) reports on four experiments. He explains the patterns of results obtained in these four studies from the principle of Verbal Politeness combined with Relevance Theory. We will come back to this point in Section 3.3 of the current article.

Table 1
Experimental conditions of the 2 × 2 design with hotel rating in stars.

Form	Content	
	Positive	Negative
Direct	"The hotel was <u>good</u>" ★★★★★	"The hotel was <u>bad</u>" ★☆☆☆☆
Indirect	"The hotel was <u>not bad</u>" ★★★★★	"The hotel was <u>not good</u>" ★☆☆☆☆

final conclusion with a fifth manipulation. Positive reviews always received a 5 star hotel rating and negative reviews had a 1 star hotel rating. No tropes, as in figurative language, irony and understatements, were used in order to communicate a clear message to the reader. An example of our stimulus material translated to English can be seen in (1). Manipulations -in this case, direct positives- are underlined. These underlined words were manipulated in the different versions of the review. Hence, while in (1) the hotel was 'cheap', it was 'not cheap', 'expensive', or 'not expensive' in the other text versions. For the original Dutch version of this text in all of its versions, we refer to Appendix A.[4]

(1) **Hotel Bella Vista**

Greece / Saronic Islands / Agkistri / Megalochori

Review by F.J. from H.

Rating: ★★★★★

Today we returned from a 10 day holiday in hotel Bella Vista on Agkistri. It was <u>a nice surprise</u>. This started at the reception desk, it was <u>pleasant</u> how our passports were dealt with. During the remainder of the week this image was confirmed time after time, really <u>good</u> service. Moreover, especially given what you get, this hotel is <u>cheap</u>. The island itself is lovely. It is a journey to get there but nature makes up for everything. Concerning the hotel I can only conclude that we did have a <u>fun</u> time here.

Written on 23/06/11.

The dependent measures consisted of nine 7-point agree-disagree scale questions per construct (hotel appreciation, text appreciation and reviewer appreciation), e.g. 'this hotel is inhospitable'; 'this review is useful', and 'this reviewer is knowledgeable', respectively. A factor analysis (Varimax) forcing three factors confirmed the existence of three distinct constructs, which together explained 59.7% of the variance. The internal consistency of these factors was good; all constructs were measured reliably (α = .91, .85 and .88, respectively).

2.1.4. Participants

A total of 168 Dutch participants filled out the questionnaire. For 8 participants, however, the data were not taken into account into the analyses e.g. because these participants had filled out the neutral answer on the vast majority of the questions. This leaves a total of 160 participants (65 men; 95 women) for Study 1, aged between 17 and 63 (*M* age = 33.0, SD = 12.7).

2.1.5. Procedure

Participants were recruited to participate in an online study on hotel reviews through email and social media (i.e. Facebook and LinkedIn). After filling out demographic details, participants were asked to imagine they were looking for information on hotels for an upcoming two-week vacation on a Greek island. Each participant then saw four reviews, one for each hotel. After each review, participants provided rates on hotel appreciation, characteristics of the writer of the review and the style of writing used in the review. Participants were also asked to indicate what they thought the aim of the study was. Finally, the participants had the opportunity to leave their email address in order to qualify for a 25 euro Bol.com gift card being raffled among participants before they were thanked for participation.

2.1.6. Analysis Study 1

To analyze the data for Study 1, three multi-level models were applied: one for text appreciation, one for reviewer appreciation and one for hotel appreciation. We opted to run models on text, reviewer and hotel appreciation as three separate dependent variables, as framing might affect evaluation of the writer, text and hotel in different ways. In the so-called fixed parts of these models, the average appreciation score is estimated separately for positive descriptions in

---

[4] All original Dutch stimulus materials can be found here http://arno.uvt.nl/show.cgi?fid=126896 from page 52 onwards.

direct form ('good'), indirect positive descriptions ('not bad'), negative descriptions in direct form ('bad') and indirect negative descriptions ('not good'). These scores are allowed to vary between items, between respondents, and due to the interaction between respondent and item. As the between-item variance and between-person variance are estimated simultaneously, we may say that a cross-classified model is in operation (see Quené and Van den Bergh, 2004). The mean appreciation scores estimated in this model can be compared in a contrast test (Goldstein, 2003; Snijders and Bosker, 1999), which yields a χ^2-distributed test statistic. For a formalization of the multi-level models used and for further explanation, we refer to Appendix B.

2.2. Results Study 1

Table 2 shows the average reviewer, text and hotel appreciation scores for the four conditions. In order to show the general pattern of results most clearly, results for reviewer and text appreciation are discussed before elaborating on hotel appreciation.

For reviewer appreciation, results indicate that authors writing positive reviews (e.g. reviews in which hotels are described as 'good'/'not bad') are evaluated more positively than authors writing negative reviews (e.g. reviews in which hotels are described as 'bad' or 'not good'; $\chi^2 = 45.21$; df = 1; $p < .001$; Cohen's $d_{\text{persons}} = 1.19$).[5] In addition, we find reviewers writing in the direct form (e.g. 'good'/'bad') to be evaluated more positively than reviewers writing in the indirect form (e.g. 'not good'/'not bad'; $\chi^2 = 31.80$; df = 1; $p < .001$; Cohen's $d_{\text{persons}} = 0.99$).

These main effects, however, should be interpreted in light of an interaction between content and form ($\chi^2 = 18.98$; df = 1; $p < .001$; Cohen's $d_{\text{persons}} = 0.75$). While for positive reviews a large difference in appreciation occurs ($\chi^2 = 52.51$; df = 1; $p < .001$; Cohen's $d_{\text{persons}} = 0.91$) – reviewers using direct forms (e.g. 'good') are evaluated more positively than reviewers using indirect forms (e.g. 'not bad') – for negative reviews, the difference between direct (e.g. 'bad') and indirect forms (e.g. 'not good') is negligible ($\chi^2 = 0.79$; df = 1; $p = .37$).

The scores for text appreciation follow a pattern similar to that for reviewer appreciation: there are main effects of both form ($\chi^2 = 34.98$; df = 1; $p < .001$; Cohen's $d_{\text{persons}} = 1.30$) and content ($\chi^2 = 57.85$; df = 1; $p < .001$; Cohen's $d_{\text{persons}} = 1.65$), and an interaction between form and content ($\chi^2 = 50.52$; df = 1; $p < .001$; Cohen's $d_{\text{persons}} = 1.23$). Again, the interaction indicates that reviews in which a positive content is communicated get higher scores when this is done using direct terminology ($\chi^2 = 132.92$; df = 1; $p < .001$; Cohen's $d_{\text{persons}} = 1.43$), whereas reviews communicating a negative content receive similar evaluations when direct and indirect forms are used ($\chi^2 = 0.11$; df = 1; $p = .74$).

While the scores for text and reviewer appreciation follow a similar pattern, a slightly different picture arises for hotel appreciation (see the third figure in Table 2). Similarly to the scores for reviewer and text appreciation, there are main effects of both content ($\chi^2 = 635.34$; df = 1; $p < .001$; Cohen's $d_{\text{persons}} = 4.44$) and form ($\chi^2 = 8.69$; df = 1; $p < .001$; Cohen's $d_{\text{persons}} = 0.53$), and these main effects are moderated by an interaction between form and content ($\chi^2 = 38.46$; df = 1; $p < .001$; Cohen's $d_{\text{persons}} = 1.09$). This interaction, however, should be specified differently than before. Readers of reviews evaluate a hotel described in positive terms even more positively, when direct terminology is used ($\chi^2 = 40.89$; df = 1; $p < .001$; Cohen's $d_{\text{persons}} = 0.77$), as was the case for text and reviewer appreciation, but this time we also find that a negative evaluation of a hotel amplifies the negative appreciation of the hotel direct when direct terminology is used ($\chi^2 = 5.42$; df = 1; $p < .001$; Cohen's $d_{\text{persons}} = 0.27$). The latter effect is small both in absolute differences and in terms of effect size, compared to the other effects observed.

2.3. Conclusion and discussion Study 1

The results of Study 1 show clear and consistent effects of content: participants respond more favorably to reviews with a generally positive evaluation. This effect extends beyond their appreciation of the hotel that is described in the review: both the text and its writer receive higher scores too. For form (direct vs. indirect), there are main effects on all three levels (reviewer, text, hotel) as well.

Each of the three dependent measures, in addition, shows interaction effects of these independent variables too. The observed interaction is partially stable across constructs: for positive reviews, reviewers, texts and hotels alike are rated higher for direct ('good') than indirect ('not bad') forms. When it comes to negative reviews, however, differences between constructs are observed. On the one hand, hotels that are described in direct terms ('bad') are deemed worse than hotels described in indirect terms ('not good'); this is the pattern predicted by the *symmetrical mitigation hypothesis*. On the other hand no such difference is found for reviewer and text appreciation, which is in line with the *positive evaluation mitigation hypothesis*.

---

[5] The effect size is often a measure that classifies the difference between conditions relative to the standard deviation (Cohen, 1988). As can be read from Table 2, there are multiple standard deviations in the multi-level models we report here. In line with earlier studies (Kamoen, 2012) we decided to calculate the effect size relative to the between-person standard deviation.

Table 2

Mean ratings for reviewer, text and hotel appreciation (scores ranging from 1–7; 7 means a positive evaluation) and the corresponding variances.

Wording			Mean	S^2_{items} [a]	$S^2_{persons}$	$S^2_{interaction}$	
Content	Form	Example					
Reviewer							
Positive	Direct	*Good*	4.86	0.22	0.29	0.85	
	Indirect	*Not bad*	4.27		0.58	0.99	
Negative	Direct	*Bad*	4.20		0.46	0.88	
	Indirect	*Not good*	4.12		0.54	0.80	
Text							
Positive	Direct	*Good*	5.44	0.20	0.40	0.91	
	Indirect	*Not bad*	4.35		0.76	1.48	
Negative	Direct	*Bad*	4.47		0.86	1.15	
	Indirect	*Not good*	4.44		0.96	1.10	
Hotel							
Positive	Direct	*Good*	5.00	0.24	0.47	0.76	
	Indirect	*Not bad*	4.45		0.53	0.81	
Negative	Direct	*Bad*	3.12		0.42	1.14	
	Indirect	*Not good*	3.31		0.45	1.15	

*Marks significance.

[a] The item variance is estimated once, this is a constraint of the model.

One question that arises from Study 1 is why a different pattern is found for hotel appreciation than for the other constructs. This may be a coincidence, as the 'extra' effect found for negative reviews here is quite small, or it may be due to a true difference between object evaluations and evaluations of reviewers and texts. In other to test this, a second study was set up, which focuses on the slightly different genre of restaurant reviews. This provides us with the opportunity to try and replicate our findings, this time for texts with a slightly different topic. If the same effects obtain, the generalizability of these for 'hotel appreciation' can be extended to 'product/service appreciation'. At the same time, we do not vary in the (more general) text type of online reviews, as these offer a realistic setting for frame manipulations of evaluative descriptions.

3. Study 2

3.1. Method

3.1.1. Aim

The main aim of Study 2 was to examine if the patterns found for hotel reviews can be extended to the genre of restaurant reviews. Therefore, the design, material and procedure for this study were similar to those in Study 1.

3.1.2. Design

As in Study 1, a 2 (form: direct versus indirect) × 2 (content: positive versus negative) design was used. This time, reviews were written about restaurants.

3.1.3. Material

Reviews were written about four fictional restaurants. Topics and wordings used were again based on a corpus study, to guarantee realistic stimulus material. Each review started with a descriptive sentence, followed by 4 sentences with one manipulation each. After an unrelated comment (e.g. about the surroundings of the restaurant), the review ended with a final conclusion containing the last manipulation. An example of a full review in the direct negative condition is provided in (2); all manipulations are underlined. As can be read, the reception was described to be 'unfriendly' in this text version. This means that, for example, the reception was termed 'friendly', 'not friendly', 'not unfriendly' in the other review versions. For the original Dutch version of this text in all of its versions, we refer to Appendix C.[6]

(2)   We received a voucher for dinner at restaurant 'De Vuurtoren' as a gift from acquaintances. It was a letdown. We had to drive quite far, so we made a reservation in advance just to be sure. Once we arrived at the restaurant, the reception was unfriendly. This impression was confirmed time and again, for instance by the personnel that was always impolite. The food and drinks were expensive, given the quality of what we got. In our opinion, this restaurant was a disappointment.

The dependent measures consisted of nine 7-point agree-disagree scale questions per construct. For text appreciation and reviewer appreciation, the same items were used as in Study 1. For restaurant appreciation, the hotel appreciation items were adapted e.g. 'I expect this restaurant serves good food'. A factor analysis (Varimax) confirmed the existence of three constructs, with together 61% variance explained. These constructs were all measured reliably (α = .89, .90 and .86 for restaurant, text and reviewer appreciation respectively).

3.1.4. Participants

A total of 121 participants filled out the questionnaire. For 8 participants, data were removed from the dataset e.g. because they had filled out the neutral answer on the majority of the questions, leaving a total of 113 participants (59 women; 54 men). The age of participants ranged from 17 to 65 (*M* age = 29.7, SD = 12.3).

3.1.5. Procedure

The procedure was similar to the procedure in Study 1, with the exception that no gift certificate was raffled.

3.2. Results

In Table 3, the average reviewer, text and restaurant appreciation scores are given. All three constructs show main effects of content and form, as well as an interaction between the two (in all cases χ^2 > 6.84; df = 1; p < .001; Cohen's $d_{persons}$ > 0.73). For reviewer appreciation, the interaction reads that in positive reviews, the direct form is preferred to the indirect form (χ^2 = 52.27; df = 1; p < .001; Cohen's $d_{persons}$ = 1.01), while for negative reviews the form chosen is irrelevant (χ^2 = 2.06; df = 1; p = .15). This is exactly the same pattern for reviewer appreciation as in Study 1 (see figure in Table 2).

The scores for text appreciation now follow the pattern that we found for hotel appreciation in Study 1. So, positive evaluations in direct terms receive higher ratings than texts in which restaurants are described as 'not bad' (χ^2 = 140.78; df = 1; p < .001 Cohen's $d_{persons}$ = 1.66). Negative content in indirect terms ('not good') is appreciated higher than negative content in direct terms ('bad'; χ^2 = 6.26; df = 1; p < .001; Cohen's $d_{persons}$ = 0.36). This latter effect is again small both in absolute sense and as compared to the other effects observed.

The scores for restaurant appreciation now follow the pattern observed for text appreciation in Study 1. Hence, a good restaurant is valued most when direct wordings are used (χ^2 = 24.17; df = 1; p < .001; Cohen's $d_{persons}$ = 0.73), whereas a bad restaurant receives equally poor evaluations irrespective of whether direct or indirect terminology is used (χ^2 = 1.00; df = 1; p = 0.32).

3.3. Conclusion and discussion Study 2

Like Study 1, Study 2 shows that the content and the form of reviews matter, and that these factors interact. When the review has a positive content, there is a consistent higher score for direct wordings ('good') over indirect forms ('not bad'). For negative reviews, no difference is observed between direct and indirect forms ('bad' and 'not good') for reviewer and product appreciation, but a small effect is found for text appreciation. When we combine Study 1 and 2, our results are generally in line with the *positive evaluation mitigation hypothesis* based on the principle of Verbal Politeness and with the results of experiment 1 and 3 in Colston (1999).

---

[6] All original Dutch stimulus materials can be found here http://arno.uvt.nl/show.cgi?fid=131056 (stimulus materials from p. 69 onwards).

Table 3

Mean ratings for reviewer, text and restaurant appreciation (scores ranging from 1–7; 7 means a positive evaluation) and the corresponding variances.

Wording			Mean	S^2_{items}[a]	$S^2_{persons}$	$S^2_{interaction}$	
Content	Form	Example					
Reviewer							
Positive	Direct	*Good*	5.05	0.34	0.44	0.79	
	Indirect	*Not bad*	4.31		0.52	1.21	
Negative	Direct	*Bad*	3.76		0.46	1.40	
	Indirect	*Not good*	3.91		0.49	0.98	
Text							
Positive	Direct	*Good*	5.50	0.33	0.54	0.71	
	Indirect	*Not bad*	4.09		0.89	1.84	
Negative	Direct	*Bad*	3.85		0.95	1.42	
	Indirect	*Not good*	4.20		0.89	1.12	
Restaurant							
Positive	Direct	*Good*	5.06	0.23	0.43	1.01	
	Indirect	*Not bad*	4.55		0.56	1.00	
Negative	Direct	*Bad*	3.41		0.48	1.31	
	Indirect	*Not good*	3.52		0.67	1.11	

*Marks significance.

[a] The item variance is estimated once, this is a constraint of the model.

However, as has been mentioned in introduction, the *positive evaluation mitigation hypothesis* can also be derived from other theoretical perspectives than Verbal Politeness. Colston (1999), for example, mentions that the pattern that we call positive evaluation mitigation can also be derived from Relevance Theory. A central tenet in Relevance Theory is the Cognitive Principle of Relevance: ''Human cognition tends to be geared to the maximization of relevance'' (Wilson and Sperber, 2004: 610). Negation is relevant in natural language use for countering an assumption part of shared background knowledge. Therefore, the meaning of direct positive, direct negative, indirect positive and indirect negative language – according to Relevance Theory – depends on the expectations shared by the interlocutors. When discourse expectations are positive, a pattern of positive evaluation mitigation may be expected, whereas in case of negative expectations the reversed pattern of negative evaluation mitigation may manifest itself.

To explain this further, we will discuss two example situations. First, imagine two people talking, A and B. Both A and B are fans of James Bond movies, and B has just seen the new Bond movie. When A sees B, he therefore asks 'How was the movie?'. As A and B both like Bond movies in general, we assume that A expects a positive evaluation. B can confirm this expectation by saying 'the movie was good', or deny it in two ways: 'the movie was bad' or 'the movie was not good'. However, when B says 'the movie was not bad', something more complex happens. The positive expectations are not really confirmed, nor denied. Therefore, the meaning of 'not bad' in such a case expresses a less positive evaluation as compared to the direct positive 'good'. This pattern is the pattern of *positive evaluation mitigation* that was also predicted by Verbal Politeness (see Fig. 2).

Second, Relevance Theory also predicts that the pattern of meaning mitigation reverses when expectations are negative, e.g. when A asks B 'How was your root canal treatment?'. In such a case, 'bad' confirms the discourse

| Good | | | | | | Bad | |
| Not bad | | | | | Not good | | |

Fig. 3. Relative distance in perceived evaluation of direct and indirect positive and negative terms when there is a negative expectation in the discourse situation (cf. Sperber and Wilson, 1986).

expectation while 'good' and 'not bad' deny it. In this case, for 'not good' something in between happens: the negative expectations are not really confirmed nor denied. Therefore, the meaning of 'not good' must be less strong than the meaning of 'bad'. A meaning pattern of *negative evaluation mitigation* as depicted in Fig. 3 is expected when the expectations in a discourse situation are negative.

Colston (1999) conducted 4 studies in total; while in Study 1 and 3 positive discourse expectations were created, in his Study 2 and 4 participants had negative expectations. His results showed that the discourse expectations indeed influenced the interpretation of direct and indirect positive and negative language use.

Our studies 1 and 2 did not contain a manipulation of a priori expectations. However, it may well be that the baseline, the default expectation, is a positive one for online reviews. Therefore, the results may be a reflection of Verbal Politeness, but they may also be explained by Relevance Theory. In order to further unravel the causes for the framing effect observed in Study 1 and 2, we therefore conducted another 2 studies in which the discourse expectation was manipulated in addition to the wording. In both of these additional studies we test two competing predictions about the cause for the framing effect we observed. First, Verbal Politeness predicts the pattern of *positive evaluation mitigation* irrespective of a priori expectations (Fig. 2). Second, Relevance Theory predicts the pattern of *positive evaluation mitigation* (Fig. 2) when discourse expectations are positive and that of *negative evaluation mitigation* (Fig. 3) when discourse expectations are negative.

4. Study 3

4.1. Method

4.1.1. Aim

The main aim of Study 3A and 3B was to examine if positive versus negative expectations about the content of the review beforehand can alter the pattern of results found in our previous two studies. Therefore, the material and procedure were similar to Study 1, with the addition of an a priori creation of expectation.

4.1.2. Design

In these studies, participants were assigned to a 2 (form: direct versus indirect) × 2 (content: positive versus negative) × 2 (expectation: positive versus negative) design.

4.1.3. Material

In Studies 3A and 3B, expectations about the hotels were created in different ways. Study 3A introduced acquaintances' (positive or negative) opinions about the hotels before showing the reviews (see Table 4 for the scenarios). In Study 3B, participants were told they were either short of money (negative expectation) or had had an unexpected financial windfall (positive expectation), which determined their selection of hotels. Both expectation

Table 4
Positive and negative expectation manipulation study 3A.

Positive expectation	Negative expectation
Say. . . Together with a good friend you are going on a last-minute holiday to the Greek island of Agkistri. Since you already booked the airline tickets, your friend has had the task to find suitable accommodation. He has selected four hotels. You've heard some <u>very positive stories</u> from your family and friends about these hotels. Apparently, the rooms are <u>clean</u> and the staff is <u>very friendly</u>. To choose between the two options, you decide to consult hotel reviews online.	Say. . . Together with a good friend you are going on a last-minute holiday to the Greek island of Agkistri. Since you already booked the airline tickets, your friend has had the task to find suitable accommodation. He has selected four hotels. You've heard some <u>very negative stories</u> from your family and friends about these hotels. Apparently, the rooms are <u>filthy</u> and the staff is <u>unfriendly at best</u>. To choose between the two options, you decide to consult hotel reviews online.

Table 5

Mean ratings for reviewer, text and hotel appreciation (scores ranging from 1–7; 7 means a positive evaluation) and the corresponding variances in study 3A.

Wording			Positive expectation			Negative expectation			
Content	Form	Example	Mean	$S^2_{persons}$	$S^2_{interaction}$	Mean	$S^2_{persons}$	$S^2_{interaction}$	S^2_{items} [a]
Reviewer									
Positive	Direct	*Good*	5.07	0.55	0.84	4.92	0.43	1.12	0.32
	Indirect	*Not bad*	4.50	0.49	1.12	4.46	0.40	0.96	
Negative	Direct	*Bad*	4.06	0.58	1.00	4.13	0.47	0.91	
	Indirect	*Not good*	4.28	0.61	0.96	4.15	0.40	1.00	

Text									
Positive	Direct	*Good*	5.61	0.59	0.77	5.50	0.78	0.70	0.23
	Indirect	*Not bad*	4.61	0.90	1.56	4.68	0.65	1.46	
Negative	Direct	*Bad*	4.23	0.95	1.32	4.52	0.74	1.14	
	Indirect	*Not good*	4.46	0.85	1.35	4.32	0.88	1.30	

Hotel									
Positive	Direct	*Good*	5.27	0.71	0.88	5.18	0.52	1.05	0.23
	Indirect	*Not bad*	4.78	0.73	1.10	4.83	0.56	0.95	
Negative	Direct	*Bad*	3.10	0.70	1.42	3.18	0.67	1.39	
	Indirect	*Not good*	3.16	0.50	1.52	3.13	0.52	1.05	

*Marks significance.

[a] The item variance is estimated once, this is a constraint of the model.

manipulations were pretested to make sure that they influenced people's attitudes. The reviews were the same as those used in Study 1, as were the questions for the dependent variables.[7]

4.1.4. Participants

A total of 241 participants filled out the questionnaire. Again, the data from 8 participants were deleted because they filled out the neutral answer on the majority of the questions, leaving a total of 233 participants (151 women; 82 men). The age of participants ranged from 18 to 67 (*M* age = 42.2, SD = 12.7).

4.1.5. Procedure

The procedure was similar to the procedure in Study 1, with the exception that participants who wrote down their email address qualified for three 10 euros Bol.com gift cards being raffled.

4.2. Results study 3A and 3B

For Study 3A and 3B the results were similar: all effects found in Study 3A also obtain for Study 3B. Therefore we only discuss the results for Study 3A (Table 5).

Results for Study 3A show that a positive or a negative expectation about the content of the review has no effect at all: the main effect of expectation, as well as the two-way and three-way interactions with expectation in it, all failed to reach significance (in all cases: $\chi^2 < 1.87$; df = 1; $p > .17$).

As was the case in Study 1 and 2, we found main effects of review content and review form, as well as an interaction between the two (in all cases $\chi^2 > 17.16$; df = 1; $p < .001$). For all three constructs (text, hotel, and reviewer appreciation), the interaction followed the pattern we saw for reviewer appreciation in Studies 1 and 2. This means that for positive reviews form matters: direct forms ('good') are evaluated more positively than indirect forms ('not bad': in all cases $\chi^2 > 57.73$; df = 1; $p < .001$; Cohen's $d_{persons} > 1.06$). For negative reviews, on the other hand, the difference between direct ('bad') and indirect forms ('not good') can be neglected (in all cases $\chi^2 < 2.89$; df = 1; $p > .09$).

4.3. Conclusion Study 3

The data in Studies 3A and 3B show main effects of content and form. The two-way interaction patterns indicate a significant difference between 'good' and 'not bad', for hotel, text and reviewer alike, but no difference between 'bad' and 'not good'. In Studies 3A and 3B, no main nor interaction effects with expectation were found. Therefore, Verbal Politeness rather than Relevance Theory seems to explain the effects found; the pattern of *positive evaluation mitigation* is observed irrespective of a priori expectations.

5. Overall conclusion and discussion

In natural language use, writers of online reviews apply direct and indirect language in order to express positive (e.g. 'good' or 'not bad') and negative (e.g. 'bad' or 'not good') evaluations. Based on Argumentative Orientation (Holleman and Pander Maat, 2009), Relevance Theory (Sperber and Wilson, 1986; Wilson and Sperber, 2004), and Verbal Politeness (Brown and Levinson, 1987), we formulated competing predictions about the effects of these frame choices. In order to test these predictions, the effects of direct and indirect positive and negative language use in online reviews were investigated in a series of four studies.

Results show that framing in online reviews affects not only the evaluation of the attitude object (in our cases: the hotel or the restaurant), but also the evaluation of the reviewer and the text, and that it does so in a rather consistent manner. More specifically: for positive reviews, direct wordings ('good') are seen as more positive than indirect wordings ('not bad'), whereas for negative reviews there is generally no difference for direct and indirect wordings ('bad' versus 'not good'). These effects are large in terms of effect size. In addition, the effects occur irrespective of whether the reader of the review has a positive or a negative expectation about the attitude object in advance.

When these results are considered in light of our three hypotheses, a first conclusion is that Relevance Theory cannot account for them. Relevance Theory states that the interpretation of evaluative descriptions in different forms is influenced by contextual assumptions such as a priori discourse expectations of the readers. Therefore, this theory predicted an interaction between the expectation about the attitude object and the framing manipulation. While pretests proved that the expectation manipulation did create different expectations, these expectations did not result in an interaction effect with the wording manipulation. Therefore, our findings are in contrast to Colston's (1999) empirical findings when it comes to the interaction

---

[7] All original Dutch materials for Study 3A and 3B can be found here http://arno.uvt.nl/show.cgi?fid=130656 (stimuli from p. 48 onwards) and here http://arno.uvt.nl/show.cgi?fid=130810 (stimuli from p. 53 onwards).

between a priori expectations and a framing manipulation. One difference between Colston's studies and the ones presented here, lies in the task participants faced. Colston asked participants what they thought a speaker meant to express, given a particular scenario, and thus, to reflect on the intended meaning of the speaker. In our studies, participants were asked, in a natural task of judging reviews, to give their own opinions about an attitude object given someone else's evaluation.

Second, results of our studies are partly in line with the *symmetrical mitigation hypothesis*, based on Argumentative Orientation (Holleman and Pander Maat, 2009). This principle states that speakers choose to produce frames that match their opinions, and listeners interpret the frame chosen as indicative of an evaluative stance taken by the speaker. As the principle of Argumentative Orientation focuses on communicating opinions, it seemed applicable to such evaluative contexts as online reviews. From this principle we derived the prediction that the meaning of indirect language use is always understood to communicate a less strong meaning as compared to a direct alternative; if the speakers were really positive or negative, they would have chosen a direct wording. Hence, this principle does account for the differences in interpretation for communicating a positive message ('good' and 'not bad'), but on its own and in its current form, it would also predict a difference in meaning for negative reviews ('bad', 'not good').

Third, and most importantly, results of the current research match with the *positive evaluation mitigation hypothesis* derived from Verbal Politeness (Brown and Levinson, 1987). Verbal Politeness assumes people want to be polite. Therefore, communicating a negative message, they use indirect terminology for doing so. However, as they use such indirect language just to be polite, the meaning of indirect negative language ('not good') is in fact just as negative as the meaning of direct negative language ('bad'). Politeness does not play a role in communicating a positive message; therefore, following Gricean maxims, the meaning of an indirect positive term ('not bad') rather than a direct positive term ('good') reflects a less positive opinion. Verbal Politeness therefore predicts that words such as 'not bad' are less positive than words such as 'good', while 'not good' has an equally negative connotation as compared to 'bad'; this prediction matches the results of our studies.

The fact that Verbal Politeness serves as an explanation for the different meanings language users assign to direct and indirect positive and negative language in online reviews, is striking. After all, the evaluation involved in online reviews is not a Face-Threatening Act (FTA): it is not an evaluation of a person, but a more abstract experience, and the people who are evaluated (perhaps the owner of the hotel, chef at a restaurant, personnel etc.) are unknown to the writer and not present. Therefore, it would be interesting to investigate this principle again in a new study, which uses more direct measures of Verbal Politeness. For example, a production study can be conducted in which participants write online reviews while they think aloud. Writers' verbalizations in those contexts can shed more light on the extent to which they take politeness into account, and whether a negative review is perceived as an FTA. Moreover, it would be interesting to investigate these issues in different languages and cultures to see if these are cultural differences in the interpretation of negation (e.g. British English use of 'not being amused').

Another suggestion for future research is to investigate the relation between AO and VP. Up to this point, Verbal Politeness has been treated as a separate entity or theory from which hypotheses can be derived. However, Colston (1999) suggests that Verbal Politeness may be seen as a general cognitive tendency that ''may or may not be affixed to other accounts'' (p. 253). AO allows for factors than the speaker's stance to affect the production and interpretation of frames (see Holleman and Pander Maat, 2009: 2208). If AO and VP were to play a role simultaneously, participants in a production study as described above would be predicted to verbalize both argumentative and politeness considerations for their framing choices.

Another way to further investigate the relation between AO and VP is to conduct studies similar to Holleman and Pander Maat (2009), but in a 2×2 design. In previous studies demonstrating the existence of AO, the frames compared reflect a comparison between only two frames: a direct positive and a direct negative frame, e.g. *say, you know that the exam was* 'easy'/'difficult', *would you say that the passing rate is 60%, or that the failure rate is 40%?*. It is hard to imagine how an effect of VP would manifest itself when only two conditions are compared. Therefore, we plan to conduct a follow-up study using the same materials as Holleman and Pander Maat (2009), but adapting the contrasts to include both direct and indirect frames, e.g. *Say, you know that the exam was* 'easy'/'difficult'/'not easy'/'not difficult', *would you say that the passing rate is 60% or that the failure rate is 40%?*. Our prediction, then, is that effects of Verbal Politeness should be visible under those conditions too.

All in all, the present studies expand on the existing framing literature by examining attribute framing in the real language use situation of online reviews, using different types of linguistic contrasts as compared to previous framing research. In the context of online reviews, readers are sensitive to rather subtle nuances in evaluations proffered by reviewers: we are *not too bad* at reading between the lines.

Appendix A. Example of Dutch stimulus material Study 1

Below one of the original Dutch stimulus materials has been depicted. The table below can be used to construct the four versions of this text.

(1) **Hotel Bella Vista**

Griekenland / Saronische eilanden / Agkistri / Megalochori

Review door F.J. te H.

Waardering: ★★★★★

Vandaag teruggekomen van 10 dagen Agkistri in hotel Bella Vista. Was __1__. Het begon al bij de balie, __2__ om te zien hoe er met onze paspoorten werd omgegaan. In de rest van de week werd dit beeld keer op keer bevestigd, echt __3__ service. Bovendien is dit hotel, zeker in vergelijking met wat je krijgt, __4__. Het eiland zelf is heerlijk. Een hele reis om er te komen maar de natuur maakt alles goed. Wat betreft het hotel kan ik niet anders dan concluderen dat we er __5__ tijd hebben gehad.

Geschreven op 23/06/11.

Manipulation number	Text version			
	Direct positive	Indirect positive	Direct negative	Indirect negative
1	een meevaller	geen tegenvaller	een tegenvaller	geen meevaller
2	prettig	niet onprettig	onprettig	niet prettig
3	goede	geen slechte	slechte	geen goede
4	goedkoop	niet duur	duur	niet goedkoop
5	een plezierige	geen onplezierige	een onplezierige	geen plezierige

Appendix B. Multi-level models

The data in our studies have a relatively complex structure. This is due to the fact that each construct is measured with multiple items (therefore, each rating is nested within a specific item, i.e. survey question) and multiple respondents answered each question (therefore, each rating is nested within a specific person at the same time). Traditional ANOVAs are not fully equipped to handle the structure of such complex data correctly, as variance due to items and the variance due to respondents cannot be estimated simultaneously. As a consequence, the total variance is usually underestimated, which results in too optimistic test statistics (Quené and Van den Bergh, 2004). To decrease the risk of type 1 errors, we used multilevel models to analyze the data in the current series of studies. Although technically complex, such multilevel models are able to handle the structure of the data correctly and hence, type 1 errors are less likely to occur.

The multi-level models in each of our studies have a similar set-up. As an example, we therefore only elaborate on the multi-level model applied in Study 1 for estimating the effect of form on text appreciation. In this model, average text appreciation scores are estimated separately for positive descriptions in direct form (e.g. *good*), positive descriptions in indirect form (e.g. *not bad*), negative descriptions in direct form (e.g. *bad*) and negative descriptions in indirect form (e.g. *not good*). These scores are allowed to vary between items, between respondents,[8] and due to the interaction between respondent and item. As the between-item variance and between-person variance are estimated simultaneously in the model, a cross-classified model is in operation (see Quené and Van den Bergh, 2004).

Eq. (A1) gives a formalization of this model. In this model, $Y_{(jk)}$ indicates the text appreciation score for participant j ($j = 1, 2, \ldots, 160$) and item k ($k = 1, 2, \ldots, 36$). In addition, there are four dummies (D), one for each of the conditions: positive descriptions in direct form (D_POS_DIR$_{(jk)}$), positive descriptions in indirect form (D_POS_INDIR$_{(jk)}$), negative descriptions in direct form (D_NEG_DIR$_{(jk)}$), and negative descriptions in indirect form (D_NEG_INDIR$_{(jk)}$). These dummies are set to 1 if the observation matches the prescribed type. Using these dummies, four average text appreciation

---

[8] In our experimental design, each respondent read one review about each of the four hotels, and each respondent was confronted with each experimental condition only once. Therefore, the between-person variance in fact equals the between-hotel variance, as in the design chosen these sources of variance cannot be separated. Therefore, the variance modeled at level 2 has in fact two interpretations; it reflects both the differences between persons and the differences between hotels. Separating these variances can only be done when one chooses an experimental design in which each participant reads multiple reviews about the same hotel and in which the same respondent is confronted with the same review version multiple times for different hotels. As this would create an unnatural task, we prefer the set-up described in Section 2.1.

scores are estimated (β_1, β_2, β_3, β_4) which are allowed to vary between items (v_{0k}), persons (u_{1j0}, u_{2j0}, u_{3j0}, u_{4j0}), and due to residual factors ($e_{1(jk)}$, $e_{2(jk)}$, $e_{3(jk)}$, $e_{4(jk)}$).

$$\text{Logit}\,(Y_{(jk)}) = \text{D\_POS\_DIR}_{(jk)}(\beta_1 + u_{1j0} + e_{1(jk)}) + \text{D\_POS\_INDIR}_{(jk)}(\beta_2 + u_{2j0} + e_{2(jk)})$$
$$+ \text{D\_NEG\_DIR}_{(jk)}\,(\beta_3 + u_{3j0} + e_{3(jk)}) + \text{D\_NEG\_INDIR}_{(jk)}(\beta_4 + u_{4j0} + e_{4(jk)}) + v_{0k}. \tag{A1}$$

The model in A1 can be described as a cross-classified model (Quené and Van den Bergh, 2008), as the model accounts for each observation to be nested within items and persons at the same time. All residuals are normally distributed with an expected value of zero, and a variance of respectively S^2_{u1j0}, u_{2j0}, S^2_{u3j0}, S^2_{u4j0} and S^2_{v0k}. Please note that, as a constraint of the model, the question variance S^2_{v0k} is estimated only once for all four conditions together.

The models used for the other dependent variables in Study 1 and all three constructs in Study 2 are equal to the one described in detail here for text appreciation in Study 1. For Studies 3A and 3B, the models are similar, but slightly more complex, as they involve an extra binary factor in the design: expectation (positive or negative). This necessitates the estimation of eight fixed parameters and their corresponding variances.

Appendix C. Example of Dutch stimulus material Study 2

Below one of the original Dutch stimulus materials has been depicted. The table below can be used to construct the four versions of this text.

Manipulation number	Text version			
	Direct positive	Indirect positive	Direct negative	Indirect negative
1	een topper	geen afknapper	een afknapper	geen topper
2	vriendelijk	niet onvriendelijk	onvriendelijk	niet vriendelijk
3	altijd beleefd	nooit onbeleefd	altijd onbeleefd	nooit beleefd
4	goedkoop	niet duur	duur	niet goedkoop
5	een succes	geen flop	een flop	geen success

References

Banks, Sara M., Salovey, Peter, Greener, Susan, Rothman, Alexander J., Moyer, Anne, Beauvais, John, Epel, Elissa, 1995. The effects of message framing on mammography utilization. Health Psychol. 14, 178–184.

Beks, Martijn, 2012. Wat is er mis met een hotel dat niet slecht is? Experimenteel onderzoek naar de effecten van positieve en negatieve framing in online hotelbeoordelingen. (What is wrong with a hotel that is not bad? Experimental research on the effects of positive and negative framing in online hotel reviews (Unpublished master's thesis). Tilburg University.

Brown, Penelope, Levinson, Stephen C., 1987. Politeness: Some Universals in Language Usage. Cambridge University Press, Cambridge.

Cohen, Jacob, 1988. Statistical Power Analysis for the Behavioral Sciences, 2nd ed. Lawrence Erlbaum Associates, New Jersey.

Colston, Herbert L., 1999. Not good'' is ''bad'', but ''not bad'' is not ''good'': an analysis of three accounts of negation asymmetry. Discourse Process. 28 (3), 237–256.

Goldstein, Harvey, 2003. Multilevel Statistical Models. Edward Arnold, London.

Gretzel, Ulrike, Yoo, Kyung H., 2008. Use and impact of online travel reviews. In: O'Connor, P., Höpken, W., Gretzel, U. (Eds.), Information and Communication Technologies in Tourism. Springer, Vienna, New York, pp. 35–46.

Grice, Herbert P., 1975. Logic and conversation. In: Cole, P., Morgan, J.L. (Eds.), Syntax and Semantics 3: Speech Acts. Academic Press, New York, pp. 41–58.

Holleman, Bregje C., Pander, Maat Henk L.W., 2009. The pragmatics of profiling: framing effects in text interpretation and text production. J. Pragmat. 41 (11), 2204–2221.

Horn, Laurence R., 1989. A Natural History of Negation. University of Chicago Press, Chicago and London.

Kahneman, Daniel, Tversky, Amos, 1979. Prospect theory: an analysis of decision under risk. Econometrica 47 (2), 263–292.

Kahneman, Daniel, Tversky, Amos, 1984. Choices, values, and frames. Am. Psychol. 39 (4), 341–350.

Kamoen, Naomi, 2012. Positive versus negative: a cognitive perspective on wording effects for contrastive questions in attitude surveys (Doctoral dissertation). LOT Dissertation Series, Utrecht.

Langacker, Ronald W., 2008. Cognitive Grammar: A Basic Introduction. University Press, Oxford.

Levin, Irwin P., Schnittjer, Sara K., Thee, Shannon L., 1988. Information framing effects in social and personal decisions. J. Exp. Soc. Psychol. 24 (6), 520–529.

Levin, Irwin P., Schneider, Sandra L., Gaeth, Gary J., 1998. All frames are not created equal: a typology and critical analysis of framing effects. Organ. Behav. Hum. Decis. Process. 76 (2), 149–188.

Litvin, Stephen W., Goldsmith, Ronald E., Pan, Bing, 2008. Electronic word-of-mouth in hospitality and tourism management. Tour. Manag. 29 (3), 458–468.

Marteau, Theresa M., 1989. Framing of information: its influence upon decisions of doctors and patients. Br. J. Soc. Psychol. 28 (1), 89–94.

McKenzie, Craig R.M., Nelson, Jonathan D., 2003. What a speaker's choice of frame reveals: reference points, frame selection and framing effects. Psychon. Bull Rev. 10 (3), 596–602.

Quené, Hugo, Van den Bergh, Huub, 2004. On multilevel modeling of data from repeated measures designs: a tutorial. Speech Commun. 43 (1–2), 103–121.

Quené, Hugo, Van den Bergh, Huub, 2008. Examples of mixed-effects modeling with crossed random effects and with binomial data. J. Mem. Lang. 59 (4), 413–442.

Sher, Shlomi, McKenzie, Craig R.M., 2006. Information leakage from logically equivalent frames. Cognition 101 (3), 467–494.

Sher, Shlomi, McKenzie, Craig R.M., 2009. Levels of information: a framing hierarchy. In: Keren, G. (Ed.), Perspectives on Framing. Psychology Press, New York.

Snijders, Tom A.B., Bosker, Roel J., 1999. Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling. Sage Publications, London.

Sperber, Dan, Wilson, Deirdre, 1986. Relevance: Communication and Cognition. Blackwell, Oxford.

Tribushinina, Elena, 2008. Cognitive Reference Points: Semantics Beyond the Prototypes in Adjectives of Space and Colour. LOT, Utrecht.

Vermeulen, IvarE., Seegers, Daphne, 2009. Tried and tested: the impact of online hotel review on consumer consideration. Tour. Manag. 30 (1), 123–127.

Wilson, Deirdre, Sperber, Dan, 2004. Relevance theory. In: Horn, L.R., Ward, G. (Eds.), The Handbook of Pragmatics. Blackwell Publishing, Malden, MA, pp. 607–632.

Ye, Qiang, Law, Rob, Gu, Bin, 2009. The impact of online user reviews on hotel room sales. Int. J. Hosp. Manag. 28 (1), 180–182.

Dr. Naomi Kamoen works as a post-doc researcher at the Utrecht Institute of Linguistics OTS at Utrecht University and as an assistant-professor at Tilburg University. Her research focuses on language use and the cognitive processes of text interpretation. Specifically, she investigates the effects of issue framing and valence framing in real language use contexts such as reviews, surveys, and voting advice applications.

Dr. Maria Mos is a linguist who teaches courses related to language and methodology at Tilburg University. Her research interests include mental representations of language and their manifestations in (meta-)linguistic behavior. She investigates language development in (older) children as well as the nature of linguistic units including collocations, fixed expressions and common phrases in different text types.

Willem F.S. Dekker (Robbin) MSc MA (b. 1989) is an alumnus from Tilburg University. After he graduated in 2011 in Social Psychology at the faculty of Social and Behavioral Sciences, he continued his career as a graduate student at the Humanities Faculty. In 2013 he graduated in Communication and Information Sciences with his research on the effects of framing in online restaurant reviews. Currently he is working as Communication Officer at the Tilburg Law & Economics Center (TILEC) of Tilburg University.