Stereotype Transmission and Maintenance Through Interpersonal Communication: The Irony Bias

Christian Burgers\(^1\) and Camiel J. Beukeboom\(^1\)

**Abstract**

In interpersonal communication, stereotypes are predominantly transmitted through language. Linguistic bias theory presupposes that speakers systematically vary their language when communicating stereotype-consistent and stereotype-inconsistent information. We investigate whether these findings can be extended to verbal irony use. The *irony bias* posits that irony is more appropriate to communicate stereotype-inconsistent than stereotype-consistent information. Three experiments support this hypothesis by showing that irony is found more appropriate (Experiments 1-2) and used more often (Experiment 3) in stereotype-inconsistent than in stereotype-consistent situations. Furthermore, linguistic biases have important communicative consequences, because they implicitly serve to maintain stereotypic expectancies. Experiment 4 shows that irony shares this characteristic with other linguistic biases, in that irony—compared to literal language—leads to more external attribution. Taken together, these results indicate that stereotypic expectancies are subtly revealed and confirmed by verbal irony, and that verbal irony plays an important role in stereotype communication and maintenance.

**Keywords**

stereotyping, interpersonal communication, linguistic bias, verbal irony, language

Social stereotypes are often transmitted to the general public by means of the mass media (e.g., Ramasubramanian, 2011; Schemer, 2012). Next to the mass media,
people learn about, form, and share societal beliefs and attitudes through interpersonal communication (e.g., Rimal, Limaye, Roberts, Brown, & Mkandawire, 2013; Seo & Matsaganis, 2013). Various authors highlight the importance of the type of language used in interpersonal communication about social stereotypes, as language can serve as a subtle cue for stereotyping (Collins & Clément, 2012; Lee, 2007). Linguistic bias theory provides a theoretical framework for how social stereotypes are communicated and maintained through language (Beukeboom, 2014; Maass, 1999; Wigboldus & Douglas, 2007).

Research on linguistic biases shows how speakers systematically vary their language use in situations in which stereotypical expectancies are confirmed or violated. For instance, speakers who communicate about situations in which their stereotypical expectancies are violated (compared to confirmed) use more concrete words (Wigboldus, Semin, & Spears, 2000), more explanations (Sekaquaptewa, Espinoza, Thompson, Vargas, & von Hippel, 2003), and more negations (Beukeboom, Finkenauer, & Wigboldus, 2010). Such systematic variations in language use have important communicative consequences: Not only do they implicitly reflect speakers’ stereotypical expectancies, but they can also strengthen these expectancies in recipients (Beukeboom et al., 2010; Wigboldus et al., 2000). For instance, news consumers experience higher levels of prejudice about minority groups after reading a news article on negative news about these groups (e.g., on the topic of migrants and crime) when these news articles follow linguistically biased communication patterns (Geschke, Sassenberg, Ruhrmann, & Sommer, 2010; see also von Hippel, Sekaquaptewa, & Vargas, 1997). As such, biased language use has been noted as one of the most important mechanisms of the transmission and maintenance of social stereotypes in interpersonal communication (Collins & Clément, 2012; Wigboldus & Douglas, 2007).

Although research revealed some systematic linguistic means that reflect and maintain stereotypic expectancies, other types of linguistic biases undoubtedly exist, yet remain undiscerned (Beukeboom, 2014). In order to identify such biases, it is first important to know which type of language is typically used in interpersonal discussions related to expectancies and prejudice. Various discourse analyses show that verbal irony seems to be used often in such discussions (cf. Korobov, 2005; Speer & Potter, 2000). Furthermore, irony has often been associated with expectancy violations (Kihara, 2005; Wilson & Sperber, 2004). This indicates that verbal irony could be used as a linguistic strategy to indirectly communicate expectancies and prejudice.

Irony allows speakers to communicate “a literal evaluation that is implicitly contrary to its intended evaluation” (Burgers, van Mulken, & Schellens, 2011, p. 190; see also Partington, 2007). For instance, when a professor makes a dim remark, a speaker could ironically comment “Gee, he is really smart!” By definition, such ironic comments communicate two meanings: a literal meaning (smart) and an intended meaning (stupid). We argue that this double meaning makes irony particularly appropriate in the case of violated stereotypic expectancies. That is, irony allows speakers to comment on the situation by an implied opposite intended meaning. Simultaneously, the literal (propositional) meaning of the ironic comment could echo the speaker’s expectancy that is violated in the current situation (e.g., professors are usually smart; cf. Wilson &
Sperber, 2004). Consequently, by using irony, senders could not only reveal their stereotype beliefs in language use, but also—in line with other linguistic biases—activate them in recipients.

In the present paper, we aim to expand linguistic bias theory and introduce the irony bias by investigating whether a violation of stereotypic expectancies induces senders to use irony to comment on a target’s behavior, and whether this facilitates the transmission of stereotypic beliefs to recipients. In order to build our argument, we first review the literature on linguistic biases and the literature on irony. Subsequently, we integrate these strands of literature to formulate our hypotheses on the irony bias, and report on four experiments.

**Expectancy Violations and Linguistic Biases**

Research on expectancy violations has shown that situations in which stereotypic expectancies are violated attract more attention and may be more persuasive than situations in which expectancies are confirmed, mainly because the former include novel information (Bergan, 2012). For instance, exposure to news messages or entertainment media with counter-stereotypical exemplars (Mastro & Tukachinsky, 2011; Ramasubramanian, 2011; Ramasubramanian & Oliver, 2007) can reduce negative prejudice toward the stereotyped group. However, such effects disappear when counter-stereotypical exemplars are perceived as atypical (Bodenhausen, Schwarz, Bless, & Wänke, 1995). In such cases, the exemplar’s counter-stereotypicality is not attributed to stable characteristics of the stereotyped group (internal attribution), but rather to individual situational circumstances (external attribution). Consequently, people do not adjust their cognitions about the stereotyped group (Bodenhausen et al., 1995).

Research on linguistic biases shows that language use plays a crucial role in discounting the persuasive effects of counter-stereotypical exemplars and observed behaviors. A linguistic bias is defined as a systematic asymmetry in word choice as a function of the social category—and its associated stereotypic expectancies—to which the target belongs (Beukeboom, 2014; Wigboldus & Douglas, 2007). In general, linguistic biases show that people vary their language use in communications about stereotype-inconsistent versus stereotype-consistent information. These variations allow speakers to subtly depict counter-stereotypic information as atypical and caused by transient situational causes, rather than stable dispositional traits. That is, the consistent pattern in known linguistic biases is that people use language that implies external (situational) attribution to talk about stereotype-inconsistent situations, and language that implies internal dispositional attribution to talk about stereotype-consistent situations (Beukeboom et al., 2010; Sekaquaptewa et al., 2003; Wigboldus et al., 2000).

For example, the linguistic expectancy bias (LEB; Maass, Ceccarelli, & Rudin, 1996; Maass, Milesi, Zabbini, & Stahlberg, 1995; Wenneker, Wigboldus, & Spears, 2005; Wigboldus et al., 2000; Wigboldus, Semin, & Spears, 2006) shows that language abstraction differs as a function of stereotype consistency. Thus, a specific behavior—like solving an advanced math problem—is more likely described in abstract terms when it is stereotype consistent with the actor’s category (e.g.,
the professor is smart), but more concretely when it stereotype inconsistent (e.g., the garbage man solved the math problem correctly). Given that abstract descriptions provide more information about the actor’s stable dispositional qualities and less information about the specific situation (Semin & Fiedler, 1988), inferences drawn from these descriptions confirm existing stereotypic beliefs. That is, the abstract words used for stereotype-consistent behaviors imply attribution to stable traits that are likely to be repeated (internal attribution), whereas the concrete words used for stereotype-inconsistent behaviors imply attribution to transient and situationally determined causes, suggesting exceptions to the rule (external attribution; Wigboldus et al., 2000). Such biased communicative behavior is most prevalent in situations in which the stereotyped target is an out-group member (Wigboldus, Spears, & Semin, 2005) and when speaker and addressee share common ground (Fiedler, Bluemke, Friese, & Hofmann, 2003).

A similar argument was made about the stereotypic explanatory bias (SEB; Sekaquaptewa et al., 2003) and negation bias (NB; Beukeboom et al., 2010). The SEB shows that descriptions of stereotype-inconsistent behavior contain more explanations aimed at clarifying the apparent inconsistency (e.g., because the garbage man probably copied the right answer) than descriptions of stereotype-consistent behavior. The NB posits that negations (e.g., not stupid, rather than smart) are used more often to describe stereotype-inconsistent compared to stereotype-consistent behaviors. Thus, in stereotype-inconsistent situations, senders are likely to reveal their opposite expectancy by using negations like the garbage man is not stupid. In descriptions of stereotype-consistent behavior, however, the use of negations is less likely, while the use of affirmations increases (e.g., the garbage man is stupid; the professor is smart). Like the LEB, the SEB and the NB have important communicative consequences: When a target person’s behavior is described with explanations or negations, recipients are likelier to attribute the behavior to transient external causes rather than stable internal traits, compared to when the same behavior is described without explanations or with affirmations.

In sum, research on linguistic biases shows that stereotype-consistent and stereotype-inconsistent information is formulated in such a way that stereotypic knowledge remains intact (Beukeboom, 2014). When expectancy-inconsistent behaviors are linguistically framed as atypical one-time events, the counter-stereotypic information will have little persuasive effect on cognitions. Consequently, biased behavior descriptions re-confirm and strengthen stereotypes, even in the face of stereotype-inconsistent information. To understand how stereotypes are communicated and maintained, it is important to reveal the different linguistic means through which this occurs (Harwood, 2010). The present paper sets out to shed more light on this phenomenon by investigating whether a selective use of verbal irony has a similar function of stereotype maintenance.

**Verbal Irony**

Verbal irony is often used in interpersonal communication: About 8% of turns in conversations between friends (Gibbs, 2000) and 7.4% of e-mails sent to friends (Whalen,
Pexman, & Gill, 2009) contain irony. No comment is ever ironic out of context; every comment can be literal or ironic depending on the context (Wallace, 2015). In case it refers to a clever performance, the comment “Gee, he is really smart” is literal and reflects the propositional meaning. In case it refers to a dim performance, it is ironic.

The irony literature shows that irony is typically used in situations where the speaker and addressee are friends and share common ground (e.g., Kreuz & Link, 2002; Pexman & Zvaigzne, 2004). Furthermore, irony usage is moderated by situational valence. This is reflected in the “asymmetry constraint,” which states that irony is generally more appropriate in negative than in positive situations (Kreuz & Link, 2002; Matthews, Hancock, & Dunham, 2006). Thus, regardless of expectations, ironic comments about someone’s negative behavior are likelier than ironic comments about positive behavior. One explanation for this asymmetry lies in the fact that, in negative—but not in positive—situations, irony can serve as a politeness strategy (Brown & Levinson, 1987; Dews & Winner, 1995). Irony enables speakers to criticize somebody without directly mentioning the negative behavior, thereby mitigating the blow.

Various other communicative goals have also been identified for using ironic comments like conveying humor (Matthews et al., 2006) and criticizing targets (Matthews et al., 2006). In some cases, irony can be perceived as particularly negative. Such “bitter and derisive statements that employ verbal irony as a device” are known as sarcasm (Kreuz, Roberts, Johnson, & Bertus, 1996, p. 87). Sarcasm is thus a specific form of irony (see also Attardo, 2000; Gibbs, 2000; Hancock, 2004, and others).

The decision to use an ironic comment is usually made in a split second, and usually without conscious planning (Gibbs, 2012). The linguistics literature suggests that well-designed ironic comments should meet two conditions: They should be both literally inappropriate and relevant to the context in which the comment is made (Attardo, 2000). That is, a comment like “What nice weather,” is literally inappropriate when it rains, yet still adds to the conversation because it communicates a relevant contrary ironic meaning.

Relevance theory (Wilson & Sperber, 2004), an influential theory in linguistics, posits that an utterance is relevant when it connects to background information speakers and/or recipients have (cognitively) available at the time of speaking. As speakers strive to make their communications optimally relevant, irony should connect to this background information and include an echo of an earlier statement or a general norm from which speakers implicitly dissociate themselves (Wilson & Sperber, 2004). In line with this reasoning, we expect that ironic comments could echo an existing stereotypic expectancy that is violated in a present situation.

The hypothesis that irony is especially suitable in situations in which existing (implicit) expectancies are violated also features in several other theoretical accounts in linguistics. Utsumi’s (2000) “implicit display theory of irony” states that verbal irony is typically used when a so-called ironic environment is created, which is when an existing expectation is violated. Kumon-Nakamura, Glucksberg, and Brown (1995) introduce the “allusional pretense theory of irony,” which argues that ironic comments typically allude to failed expectations. Finally, Kihara (2005) approaches irony from the perspective of mental space theory and defines irony as an allusion to failed
expectations with two distinct mental spaces: a mental space of expectation and an initial reality space that does not live up to the expectation. These linguistic theories share the assumption that irony includes a reference to existing expectancies, and that situations in which expectancies are violated invite speakers to use irony. Thus, an ironic comment like *Wow, he is really smart* allows a speaker to both note the existing expectation that the target would be smart and the actual situation in which he makes a dim remark.

Furthermore, if ironic comments indeed follow from a speaker’s stereotypic expectancies, then irony could subtly communicate these expectancies to recipients. Neurological evidence suggests that, in order to comprehend irony, recipients rely on their theory of mind abilities and form a mental image of what the speaker aims to convey (Bohrn, Altmann, & Jacobs, 2012; Spotorno, Koun, Prado, Van Der Henst, & Noveck, 2012). In this comprehension process, both the literal and ironic meaning remain active in working memory during and after processing (e.g., Akimoto, Miyazawa, & Muramoto, 2012; Giora, Fein, & Schwartz, 1998). Thus, when violated expectancies are communicated through irony, recipients are made aware of both the existing expectancy and the current violation of that expectancy.

In this respect, irony processing is similar to negation processing. Like irony, negations (e.g., *not stupid*) communicate two meanings that are both activated. That is, the negated concept (e.g., *stupid*) is comprehended first, before attaching the negation marker (*not*) that implies an opposite meaning (cf. Giora, Fein, Aschkenazi, & Alkabets-Zlozover, 2007; Mayo, Schul, & Burnstein, 2004). As a result, addressees can remember the opposite of the intended meaning (*stupid*) when they have actually heard that the target was *not stupid*. These similarities in irony and negation processing have led some scholars to argue that irony is a form of “indirect negation”; in irony, the valence of the literal evaluation is negated indirectly because irony lacks an explicit negation marker like “no” or “not” (Giora, 1995; Giora et al., 1998). The facts that negations have been shown to reflect and maintain stereotypes (Beukeboom et al., 2010) and that irony processing seems similar to negation processing provide additional support for our hypothesis that irony has a similar function to negations in stereotype maintenance.

Based on the above, we argue that irony can follow from as well as implicitly communicate stereotypic expectancies. Irony allows speakers to implicitly allude to (Kumon-Nakamura et al., 1995) or “echo” (Wilson & Sperber, 2004) existing stereotypic expectancies. By using ironic comments, speakers can refer to an existing stereotypic expectancy connected to the target of the irony through the propositional meaning and, simultaneously, by means of the intended meaning, indicate that this expectancy is violated in the current situation, thus framing the situation as atypical.

In sum, the irony bias equates to the following predictions. From a production perspective, we argue that speakers are likelier to use irony to comment on stereotype-inconsistent behavior than on stereotype-consistent behavior. Our first three experiments focus on the production side of the irony bias, and are introduced in the next section.
From a recipient perspective, the irony bias posits that a biased use of irony has important communicative consequences because recipients draw inferences from irony usage. We argue that recipients process the literal meaning of an ironic comment as an allusion to a violated (stereotypic) expectancy. Consequently, irony activates this echoed stereotypic expectancy in recipients, making recipients perceive the target situation as an exception to the rule, and a one-time event, resulting in external attribution. Our fourth experiment focuses on the recipient side and is further introduced later.

**Experiments 1, 2, and 3: Using Irony**

The goal of Experiments 1, 2, and 3 was to test whether irony is used more often in communications about stereotype-inconsistent than stereotype-consistent behavior. Across these experiments, participants were presented with positive or negative behavior that was either stereotype consistent or stereotype inconsistent for a specific actor. Subsequently, they were presented with two comments about this behavior: a literal and an ironic comment.

In Experiment 1, we asked participants to indicate how appropriate they considered the literal and the ironic utterance for the given situation. Given that irony is characterized as relevant inappropriateness (Attardo, 2000), we expect that, by definition, ironic comments should be perceived as less appropriate than literal comments. In line with our arguments on the irony bias, our main prediction is an interaction between type of comment (literal vs. ironic) and stereotype consistency (consistent vs. inconsistent). We hypothesize,

**Hypothesis 1 (H1):** Irony is rated as more appropriate to comment on stereotype-inconsistent than stereotype-consistent behavior. For literal comments, this is reversed.

Finally, because the asymmetry constraint proposes that irony is used more often in negative than in positive situations (Kreuz & Link, 2002; Matthews et al., 2006), we included the factor of behavioral valence as a control. The irony bias, however, does not assume an interaction with valence.

**Experiment 1**

**Method**

**Participants and design.** A total of 104 respondents participated in this online experiment with a $2 \times 2 \times 2$ mixed experimental design. Type of comment was a within-subjects variable, and stereotype consistency and behavioral valence were between-subjects variables. Participants were recruited via different social media (e.g., Facebook, LinkedIn). The average age of participants was 26.00 years ($SD = 11.99$, range $= 15-79$). Of all participants, 64 were female and
40 were male. Participants did not receive financial compensation for participation.

Materials, instrumentation, and procedure. Participants were presented with a situation in which an actor showed behavior that was either consistent or inconsistent with gender stereotypes and either positive or negative and asked to imagine that they would comment on the situation to a close friend. We chose this context of the good friend in this and subsequent experiments because high common ground facilitates both communication following linguistic biases (Fiedler et al., 2003) and communication with ironic comments (Kreuz & Link, 2002; Pexman & Zvaigzne, 2004). Participants were exposed to a brief scenario and photos of a person who operated a computer. In the stereotype-consistent situation, this was a man who successfully installed his new computer (positive behavior) or a woman who failed to install her new computer (negative behavior). In the stereotype-inconsistent situations, the same man and woman showed the negative and positive behavior, respectively.

After seeing the situation, participants were presented with 12 potential comments (see online Appendix A), of which six were literal and six ironic. In this and subsequent studies, literal comments reflected the propositional meaning. Comments were either positive (e.g., this person really understands how computers work) or negative (e.g., this person is clueless as to how computers work). In the situation with positive behavior, the negative comments were ironic and the positive ones were literal. In the situation with negative behavior, this was reversed. Online Appendix A gives an overview of all comments used in Experiment 1.1

Participants were subsequently asked to rate each comment on appropriateness on 7-point semantic differential scales ranging from 1 = very inappropriate to 7 = very appropriate (Cronbach’s α_literally_positive_comments = .89, α_literally_negative_comments = .91). We also included the expectedness of the situation as a control. After having chosen the comments they would use, participants were exposed to the situations again. Participants were then asked to rate the expectedness of the behavior for the specific actor (e.g., How expected is it that the person on the photo showed this behavior?) on a 7-point scale ranging from 1 = very unexpected to 7 = very expected.

Results

Expectedness. We ran an independent samples t-test to analyze whether the stereotype-consistent situations were more expected than the stereotype-inconsistent situations. Results demonstrate that the behavior was perceived as more expected in the stereotype-consistent condition (M = 5.35, SD = 1.25) than in the stereotype-inconsistent condition, M = 2.98, SD = 1.82, t(64.20) = 7.30, p < .001, r = .67, indicating that our manipulation of stereotype consistency was successful.2

Appropriateness of literal and ironic comments. To test H1, we conducted a 2 (type of comment: literal vs. ironic) × 2 (stereotype consistency: consistent vs. inconsistent) × 2 (behavioral valence: positive vs. negative) mixed analysis of variance (ANOVA) with type of comment as a within-subjects variable, stereotype consistency and
behavioral valence as between-subjects variables, and the appropriateness of the comment as a dependent variable. Table 1 shows descriptive information.

The analysis showed a main effect of type of comment, indicating that participants found the literal comments (M = 4.52, SD = 1.32) more appropriate than the ironic comments, M = 3.47, SD = 1.31, F(1, 100) = 51.96, p < .001, ηp² = .34. This main effect was qualified by an interaction effect of type of comment and stereotype consistency, F(1, 100) = 4.11, p = .045, ηp² = .04. Pairwise comparisons with Bonferroni corrections showed no differences in the appropriateness of literal comments in stereotype-consistent and stereotype-inconsistent situations (p = .78). In contrast, we found a trend indicating that ironic comments were considered more appropriate in the stereotype-inconsistent (M = 3.75, SD = 1.34) than in the stereotype-consistent situations (M = 3.29, SD = 1.59, p = .069). These findings confirm our expectations regarding the appropriateness of irony (H1).

Furthermore, we also observed an interaction between behavioral valence and type of comment, F(1, 100) = 19.39, p < .001, ηp² = .16. Pairwise comparisons with Bonferroni corrections show no differences in the appropriateness of literal comments referring to positive and negative behaviors (p = .14). Ironic comments, however, were considered more appropriate for negative (M = 3.96, SD = 1.42) than for positive behavior (M = 2.94, SD = 1.43, p < .01), which is in line with the asymmetry constraint that irony is used more often when referring to negative than to positive behavior.

Finally, our findings were qualified by a three-way interaction of type of comment, stereotype consistency and behavioral valence, F(1, 100) = 4.54, p = .036, ηp² = .04. Figure 1 shows that the perceived appropriateness of the literal comments did not differ as a function of either stereotype consistency or behavioral valence. In contrast, ironic comments were considered more appropriate in stereotype-inconsistent than in stereotype-consistent situations, but only when referring to positive (and not to negative) behavior.

### Table 1. Mean Scores (and Standard Deviations) of Appropriateness of Literal and Ironic Comments (Experiments 1-2) and Percentage of Irony (Experiment 3) as a Function of Behavioral Valence (Negative vs. Positive) and Stereotypic Consistency (Consistent vs. Inconsistent).

<table>
<thead>
<tr>
<th></th>
<th>Negative situation</th>
<th>Positive situation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Consistent</td>
<td>Inconsistent</td>
</tr>
<tr>
<td><strong>Experiment 1</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Appropriateness literal</td>
<td>4.48 (1.26)</td>
<td>4.17 (1.42)</td>
</tr>
<tr>
<td>Appropriateness irony</td>
<td>4.10 (1.46)</td>
<td>3.75 (1.36)</td>
</tr>
<tr>
<td><strong>Experiment 2</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Appropriateness literal</td>
<td>5.45 (1.06)</td>
<td>4.81 (1.13)</td>
</tr>
<tr>
<td>Appropriateness irony</td>
<td>2.73 (1.41)</td>
<td>2.92 (1.20)</td>
</tr>
<tr>
<td><strong>Experiment 3</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Percentage irony chosen</td>
<td>16.00 (28.46)</td>
<td>24.50 (32.15)</td>
</tr>
</tbody>
</table>

Note. Appropriateness is measured on 7-point scales with higher numbers indicating higher appropriateness. Percentage of irony is measured on a scale from 0% (no ironic comments chosen over literal comments) to 100% (all ironic comments chosen over literal comments). Means of variables with different subscripts within negative (a, b) and positive (x, y) behavioral situation conditions are significantly different according to pairwise comparisons with a certainty of at least p < .05. Different subscripts qualified with the symbol # indicate a trend (p < .10).
Discussion

Experiment 1 provides the first support for our hypothesis that ironic comments are more appropriate in stereotype-inconsistent than in stereotype-consistent situations, which supports the irony bias. We find empirical evidence that the irony bias can be stronger when referring to positive than to negative behaviors, which could be explained by the differential role of politeness in uttering these comments. After all, irony allows the speaker to indirectly comment on negative behavior using positive terms (Brown & Levinson, 1987). Given that politeness is less of an issue in positive situations, politeness considerations are unlikely to interfere with the effects of stereotype expectancies, making the irony bias more pronounced.

In Experiment 1, we used only one stimulus scenario. In order to test the robustness of the findings across various stimulus sets (Jackson, O’Keefe, & Jacobs, 1988), Experiment 2 was designed to conceptually replicate Experiment 1 using a different methodology and multiple sets of stimuli.

Experiment 2

Method

Participants, design, and procedure. A total of 58 respondents participated in this online experiment with a 2 (type of comment: literal vs. ironic) × 2 (stereotype consistency: consistent vs. inconsistent) × 2 (behavioral valence: positive vs. negative) × 4 (stimulus set) within-subjects experimental design. Participants were recruited via different...
social media (e.g., Facebook, LinkedIn) and directed to our online experiment. The average age of participants was 30.48 years ($SD = 13.25$, range = 18-59). Of all participants, 39 were female and 19 were male. Participants did not receive financial compensation for participation.

**Procedure, materials, and instrumentation.** The introduction screen of the online questionnaire informed participants that they participated in a study on communication. Participants were told that they would be presented with a specific situation on which they were asked to comment. They were subsequently presented with the situations in the various experimental conditions.

We created four sets of stimuli that allowed for ironic and literal comments in stereotype-consistent and stereotype-inconsistent situations. Each set consisted of two actors and two contrasting positive and negative behavioral situations. The first situation in each set was stereotype consistent for Actor 1, but stereotype inconsistent for Actor 2. For the second situation, this was reversed for the two actors. For instance, one set consisted of the situations *horribly failing a math exam* (i.e., F) and *passing a math exam with flying colors* (i.e., A+) and two actors: a blonde and a geek. In the situation in which somebody failed the math exam, the stereotype-consistent actor was the blonde and the stereotype-inconsistent actor was the geek. In the positive situation, this was reversed (i.e., the geek was the stereotype-consistent actor and the blonde was the stereotype-inconsistent actor). In order to activate stereotypic expectancies, we presented participants with photos of the actor (e.g., blonde or geek) and of the behavior (e.g., a math exam marked as F or A+). These photos were pre-tested to make sure that they accurately reflected the stereotypic categories and behaviors. The text accompanying these photos was neutral (e.g., *The person on the photo scored this result on a math exam. Which of the following would you use to comment on this situation to a good friend?*).

Each actor-situation combination was subsequently presented with two relevant comments referring to the situation. One of these was literal and the other was ironic. Note that the literal comments in the negative situation (e.g., *Of course, an incredible dumbass has been working here!* ) were the ironic comments in the positive situation. Likewise, the literal comments in the positive situation (e.g., *Of course, an incredible genius has been working here!* ) were ironic in the negative situation. We varied the presentation order such that a random actor-situation combination from Set 1 was followed by a random combination from Set 2, which was then followed by a random combination from Set 3, until all combinations from the four sets were presented. In total, participants were thus presented with 16 comment pairs referring to the actor-situation combinations. See online Appendix B for an overview of all stimuli. After seeing each situation, participants were again asked to rate the appropriateness of each comment in a conversation with a good friend on a 7-point semantic differential scale ranging from 1 = very inappropriate to 7 = very appropriate. After exposure to all situations, participants were exposed to the situations again. Participants were then asked to rate the expectedness of the behavior for the specific actor (e.g., *How expected is it that the person on the photo scored this result on a math exam?*) on a 7-point scale ranging from 1 = very unexpected to 7 = very expected.
Results

Appropriateness of literal and ironic comments. In order to test H1, we conducted a 2 (type of comment: literal vs. ironic) × 2 (stereotype consistency: consistent vs. inconsistent) × 2 (behavioral valence: positive vs. negative) repeated-measures ANOVA with the appropriateness of the comment as the dependent variable. Table 1 shows descriptive information. The analysis showed a main effect of type of comment, indicating that participants found literal comments ($M = 5.22, SD = 0.79$) more appropriate than ironic comments, $M = 2.66, SD = 1.00$, $F(1, 57) = 179.36, p < .001, \eta_p^2 = .76$. In line with our hypothesis, this main effect was qualified by an interaction effect of type of comment and stereotype consistency, $F(1, 57) = 40.19, p < .001, \eta_p^2 = .41$. Pairwise comparisons with Bonferroni corrections demonstrated that ironic comments were considered more appropriate for the stereotype-inconsistent ($M = 2.85, SD = 0.96$) than for the stereotype-consistent behaviors ($M = 2.46, SD = 1.14, p < .001$). In contrast, literal comments were considered more appropriate ($M = 5.55, SD = 0.91$) in stereotype-consistent than in stereotype-inconsistent situations ($M = 4.90, SD = 0.83, p < .001$). These findings confirm the irony bias.

We also observed a significant interaction between type of comment and behavioral valence, $F(1, 57) = 8.34, p = .005, \eta_p^2 = .13$. Pairwise comparisons with Bonferroni corrections showed no differences in the appropriateness of literal comments in positive and negative behaviors ($p = .10$). Ironic comments, however, were considered more appropriate for negative ($M = 2.83, SD = 1.24$) than for positive behaviors ($M = 2.48, SD = 0.91, p < .01$), confirming the asymmetry constraint that irony is used more often to refer to negative than to positive behavior (Kreuz & Link, 2002; Matthews et al., 2006).

Our findings were qualified by a three-way interaction of type of comment, stereotype consistency, and behavioral valence, $F(1, 57) = 5.53, p = .022, \eta_p^2 = .09$. Pairwise comparisons with Bonferroni corrections showed that literal comments were considered more appropriate in stereotype-consistent than in stereotype-inconsistent situations, regardless of behavioral valence (both $p$ values < .001). Ironic comments were considered more appropriate in positive stereotype-inconsistent situations than in positive stereotype-consistent situations ($p < .001$). For the negative behaviors, these differences were in the similar direction, but the effect was only a trend ($p = .079$). These findings suggest that the irony bias may be stronger for positive than for negative behavior (see Figure 2).

Mediation of expectedness. In order to analyze whether the perceived appropriateness of irony is indeed driven by stereotypic expectedness of the situation, we conducted a mediation analysis. We used a method for mediation analysis as outlined by Judd, Kenny, and McClelland (2001), who posit that mediation in within-subjects designs can be established if two conditions are met. First, the independent variable of stereotype consistency has to influence the proposed mediator of expectedness. We established that the stereotype-consistent situations were perceived as more expected ($M = 5.49, SD = 0.72$) than the stereotype-inconsistent situations, $M = 2.83$, ...
SD = 0.77, t(57) = 14.97, p < .001, r = .89, which means that the first condition for mediation is met. The second condition is that the dependent variable (irony appropriateness) has to be predicted by the proposed mediator (expectedness).

To obtain a single dependent measure, we calculated irony appropriateness separately for (1) the stereotype-inconsistent and (2) stereotype-consistent behaviors. Next, we computed the difference in irony appropriateness scores between conditions (irony appropriateness in stereotype-inconsistent situations − irony appropriateness in stereotype-consistent situations). This variable was regressed on two predictors: the sum of each participant’s expectedness ratings of stereotype-inconsistent and stereotype-consistent behaviors (expectedness of inconsistent behaviors + expectedness of consistent behaviors) and the difference in each participant’s expectedness ratings of inconsistent and consistent behaviors (expectedness of inconsistent behaviors − expectedness of consistent behaviors; see Judd et al., 2001).

Judd et al. (2001) argue that, if the difference scores are a significant predictor, this implies a mediation effect. The analysis showed a trend, β = −.25, t(55) = 1.92, p = .060, indicating that the effect of stereotype consistency on the appropriateness of ironic comments is driven by perceived expectedness. Furthermore, we centered the variables in the model and estimated the same model again. If the constant in the analysis with mean-centered data is not statistically significant, this indicates full mediation. In case the constant is significant, this indicates partial mediation (Judd et al., 2001). As the constant was non-significant, t(55) = 0.16, p = .87, this analysis indicates that expectedness fully mediated the effect of stereotype consistency on irony appropriateness, again confirming the hypothesis of the irony bias.

Figure 2. Three-way interaction effect between type of comment (literal vs. ironic), behavioral valence (negative vs. positive), and stereotype consistency (consistent vs. inconsistent) on the appropriateness of the comment (Experiment 2).

Note. Error bars indicate 95% CIs. CI = confidence interval.
We subsequently conducted a similar mediation analysis with the appropriateness of literal comments as the dependent variable. We also established that the effect of stereotype consistency on the appropriateness of literal comments is mediated by perceived expectedness, $\beta = .27$, $t(55) = 2.07$, $p = .043$. The constant in the mean-centered dataset was again non-significant, $t(55) = 0.20$, $p = .85$, showing full mediation. Please also note that the mediation of expectedness on the appropriateness of literal comments is in the opposite direction of the mediation of expectedness on the appropriateness of ironic comments. Both mediation analyses confirm the hypothesis of the irony bias.

**Discussion**

Experiment 2 replicates the results of Experiment 1 in showing that irony is more appropriate to describe stereotype-inconsistent behavior than stereotype-consistent behavior (H1). For literal language, this pattern is reversed. Furthermore, Experiment 2 also extends the results of the first experiment by demonstrating that the effect of stereotype inconsistency on irony appropriateness is driven by perceived unexpectedness. This indicates that, in line with our argument, irony is perceived as more appropriate when a stereotypic expectancy is perceived to be violated.

In Experiments 1 and 2, participants assessed the appropriateness of both the literal and the ironic comments. Another conventional method to measure participants’ preference is to have them choose one of two options. Such a forced-choice paradigm may correspond more closely to actual interpersonal communication in which speakers usually formulate just one comment—either literal or ironic. In order to test whether the findings from Experiments 1 and 2 replicate using this new experimental set-up, we conducted Experiment 3. We hypothesize,

**Hypothesis 2 (H2):** Speakers choose to use more irony in stereotype-inconsistent than in stereotype-consistent situations.

**Experiment 3**

**Method**

*Participants and design.* A total of 50 respondents participated in this 15-minute online experiment with a 2 (stereotype consistency: consistent vs. inconsistent) × 2 (behavioral valence: positive vs. negative) × 4 (stimulus set) within-subjects experimental design. Participants were recruited via different social media (e.g., Facebook, LinkedIn) and did not receive financial compensation for participation. The average age of participants was 24.22 years ($SD = 7.82$, range = 17-53). Of all participants, 27 were female and 23 were male.

*Materials and instrumentation.* We used the same stimulus sets and target comments as those used in Experiment 2 (see online Appendix B). However, instead of asking
participants to rate the appropriateness of each individual comment, we asked them to indicate which of the two comments they were most likely to use in interpersonal communication with a good friend. As our dependent variable, we calculated the percentage of times participants chose the ironic comment over the literal comment per condition, which could vary between 0% (never chose irony over literal comments) and 100% (always chose irony over literal comments). All other items were identical to those used in Experiment 2.

Results

Choosing irony. In general, our participants favored literal over ironic comments. They chose the ironic over the literal comments in 15.8% of cases ($SD = 18.3\%$, range $= 0\%-81\%$). Table 1 shows descriptive statistics per condition. To test our hypotheses, we conducted a $2 \times 2$ repeated-measures ANOVA with the percentage of ironic comments chosen by participants as the dependent variable. In line with H2, we found a main effect of stereotype consistency, indicating that irony was chosen more often in the stereotype-inconsistent ($M = 20.75\%$, $SD = 22.53$) than in the stereotype-consistent situations, $M = 10.75\%$, $SD = 18.90$, $F(1, 49) = 12.65$, $p = .001$, $\eta^2_p = .21$.

The analysis also showed a main effect of behavioral valence, indicating that participants choose more irony when referring to negative behavior ($M = 20.25\%$, $SD = 28.11$) than when referring to positive behavior, $M = 11.25\%$, $SD = 15.41$, $F(1, 49) = 5.62$, $p = .022$, $\eta^2_p = .10$, which confirms the asymmetry constraint (Kreuz & Link, 2002). The interaction between stereotype consistency and valence was non-significant, $F(1, 49) = 0.56$, $p = .46$.

Mediation of expectedness. In order to analyze whether the choice for ironic or literal comments is indeed driven by stereotypic expectedness of the situation, as hypothesized, we conducted a mediation analysis following the method of Judd et al. (2001) in a similar way as in Experiment 2. First, we established that the stereotype-consistent situations were perceived as more expected ($M = 5.75$, $SD = 0.69$) than the stereotype-inconsistent situations, $M = 2.57$, $SD = 0.72$, $t(49) = 16.34$, $p < .001$, $r = .92$. Next, we regressed the sum score and the difference score of each participant’s expectedness ratings of stereotype-inconsistent and stereotype-consistent behaviors on the differences in irony choice between conditions.

As the difference score was significant, mediation was established, $\beta = -.32$, $t(47) = 2.31$, $p = .025$. Furthermore, we centered the variables in the model and estimated the same model again. As the constant was non-significant ($\beta(47) = 0.68$, $p = .50$), this analysis indicates that expectedness fully mediated the choice for irony over literal comments (Judd et al., 2001), again confirming the hypothesis of the irony bias.

Discussion

Like Experiments 1 and 2, Experiment 3 provides support for our hypothesis that stereotype expectancy drives irony usage. The level of irony use in the
stereotype-consistent situations is in line with previous findings showing that irony is used in 8% of turns in conversations with friends (Gibbs, 2000). In the stereotype-inconsistent situations, however, the choice for irony almost doubled. Even though our experimental set-up is of course different from regular discourse usage, this figure demonstrates the importance of stereotype consistency in driving irony usage. Like in Experiment 2, we find that the effect of stereotype consistency on irony usage is mediated by perceived expectedness.

Experiments 1, 2, and 3 provided empirical evidence for our hypothesis on the production side of the irony bias that irony use follows from a speaker’s stereotype expectancies. In line with other linguistic biases, we also argue that the biased use of irony has communicative consequences by transmitting stereotypic expectancies to recipients. In Experiment 4, we investigate the inferences that recipients draw from ironic versus literal comments about the target of the comment.

**Experiment 4: Receiving Irony**

The previous three studies showed that a sender’s stereotypic expectancies about a target surface in a biased use of ironic versus literal comments about the target’s behavior. Importantly, and in line with other linguistic biases, we argue that the irony bias functions to implicitly communicate these expectancies to message recipients, and thus to contribute to the transmission and maintenance of socially shared stereotypes. Research on linguistic biases has demonstrated a systematic and comparable pattern of recipient inferences, which are usually measured in the types of attributions that recipients draw from biased target-behavior descriptions (Beukeboom, 2014).

Specifically, the language used for stereotype-consistent behaviors, that is, abstract language (Wigboldus et al., 2000), without explanations (Sekaquaptewa et al., 2003) or with affirmations (Beukeboom et al., 2010), tends to induce internal attributions in recipients, inferring that the described behavior is due to and caused by stable dispositional traits of the target person. Furthermore, information described with abstract language is typically generalized to the target’s group as a whole (Assilaméhou, Lepastourel, & Testé, 2013). In contrast, the language used for stereotype-inconsistent behaviors, that is, concrete language (Wigboldus et al., 2000), with explanations (Sekaquaptewa et al., 2003) or with negations (Beukeboom et al., 2010), tends to induce external attributions, inferring that the described behavior is due to transient situational circumstances that are unlikely to be repeated. Linguistic biases thus serve to maintain stereotypic expectancies on an interpersonal level, because stereotype-inconsistent behavior is linguistically framed as an exception, and an atypical one-time event, thus not changing the stereotypic expectancy regarding the social category to which the target belongs (cf. Bodenhausen et al., 1995). Through such inferences typically drawn from linguistic biases, stereotypic expectancies about social groups are perpetuated (Geschke et al., 2010).

In sum, different linguistic biases are comparable in the induced pattern of recipient inferences regarding the attribution of the target’s behavior. We expect similar recipient inferences from ironic (vs. literal) comments about behavior. The irony bias
presumes that violated stereotypic expectancies are likely communicated by irony. Given that the implicit prior expectancy communicated in the comment’s literal meaning is opposite to the current situation, the recipient is also likely to infer that the current situation is unexpected and an exception to the rule. Thus, we expect,

**Hypothesis 3 (H3):** Recipients infer more external attribution from irony as compared to literal comments with respect to the target’s behavior.

**Method**

*Participants and design.* A total of 87 respondents participated in this online experiment with a 2 (type of comment: literal vs. ironic) × 2 (behavioral valence: positive vs. negative) × 4 (stimulus set) mixed experimental design. Type of comment and behavioral valence were within-subjects variables and stimulus set was a between-subjects variable. We varied the stimuli in such a way that participants saw every stimulus set once and were thus exposed to four different stimulus sets. We also made sure that participants saw both types of comments (literal vs. ironic) once for both types of behavioral valence (positive vs. negative). So, if a participant for instance saw an ironic comment accompanying positive behavior in Stimulus Set 1, this participant would receive another type of comment–behavioral valence pairing for Stimulus Sets 2, 3, and 4. Thus, participants were only presented with each stimulus set once, but saw all experimental conditions of the experiments (negative situation–literal comment, negative situation–ironic comment, positive situation–literal comment, positive situation–ironic comment).

Like in Experiments 1, 2, and 3, we recruited participants via different social media (e.g., Facebook, LinkedIn). The average age of participants was 26.51 years (SD = 11.60, range = 18-65). Sixty-one participants were female and 26 participants were male. Participants were given the opportunity to fill out their names and e-mail addresses to enter a raffle to win a gift card.

*Materials.* Participants read that we were interested in what they infer from behavior descriptions, and that they would be presented with various situations in which a friend described the performance or behavior of another person.

Next, participants received four brief written scenarios in which a good friend or partner makes an evaluative comment about someone else’s behavior. For instance, in one of the stimulus sets, somebody wants to drive their drunk friends home. In the condition with negative behavioral valence, this person is described as being even more wasted than the drunk friends and in the condition with positive behavioral valence, this person is described as completely sober. After the behavior description, participants were presented with the comment about the target’s behavior, which was either literally negative (e.g., *Well, well, that is undoubtedly the worst designated driver ever*) or literally positive (*Well, well, that is undoubtedly the best designated driver ever*). As irony is context dependent (Wallace, 2015), the valence of the described behavior determines whether a comment is literal or ironic. Negative
Table 2. Mean Scores (and Standard Deviations) of Recipient Inferences as a Function of Situational Valence and Type of Comment (Experiment 4).

<table>
<thead>
<tr>
<th></th>
<th>Negative situation</th>
<th>Positive situation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Literal comment</td>
<td>Ironic comment</td>
</tr>
<tr>
<td>Repetition likelihood</td>
<td>59.75 (22.73)</td>
<td>62.92 (23.41)</td>
</tr>
<tr>
<td>Dispositionality</td>
<td>4.30 (1.35)</td>
<td>4.31 (1.39)</td>
</tr>
<tr>
<td>Generalizability</td>
<td>3.22 (1.60)</td>
<td>2.89 (1.59)</td>
</tr>
<tr>
<td></td>
<td>75.09 (19.34)*#</td>
<td>69.56 (23.17)*#</td>
</tr>
<tr>
<td></td>
<td>4.76 (1.38)</td>
<td>4.10 (1.68)*#</td>
</tr>
<tr>
<td></td>
<td>5.31 (1.16)</td>
<td>4.99 (1.57)</td>
</tr>
</tbody>
</table>

Note. Repetition likelihood is measured on a scale from 0 (low repetition likelihood) to 100 (high repetition likelihood). Dispositionality and generalizability are measured on 7-point scales with higher numbers indicating higher dispositionality and generalizability. Means of variables with different subscripts within negative (a, b) and positive (x, y) behavioral situation conditions are significantly different according to pairwise comparisons with a certainty of at least $p < .05$. Different subscripts qualified with the symbol # indicate a trend ($p < .10$). For generalizability, we only found two main effects indicating that generalizability was higher for literal comments and positive situations than ironic comments and negative situations, respectively.

Results

For each dependent variable, we computed mean ratings over the four stimulus sets across experimental conditions. These were analyzed with 2 (type of comment: literal vs. ironic) × 2 (behavioral valence: positive vs. negative) repeated-measures ANOVAs. Table 2 shows descriptive statistics.

For repetition likelihood, we found no main effect of type of comment, $F(1, 86) = 0.37, p = .54$. We did find a main effect of behavioral valence, $F(1, 86) = 16.59, p < .001, \eta_p^2 = .16$, indicating that participants estimated the repetition likelihood of positive behaviors ($M = 72.33\%$, $SD = 16.88$) higher than that of negative behaviors ($M = 61.33\%, SD = 19.22$). We also observed an interaction effect between behavioral valence and type of comment, $F(1, 86) = 4.59, p = .03, \eta_p^2 = .05$. Positive behaviors were rated as more likely to be repeated after literal comments ($M = 75.09\%$, $SD = 19.34\%$) than after ironic comments ($M = 69.56\%$, $SD = 23.17\%$), whereas the opposite was true for negative behaviors ($M = 59.75\%$, $SD = 22.73\%$ vs. $M = 62.92\%$, $SD = 23.41\%)$.
and type of comment, \( F(1, 86) = 4.80, p = .031, \eta_p^2 = .05 \). Pairwise comparisons with Bonferroni corrections showed no differences between comment types in the situation with negative behavior, \( p = .25 \). In the situations with positive behavior, we observe, in line with predictions, that repetition likelihood is lower for ironic than for literal comments, \( p = .052 \).

For dispositionality, we observed a main effect of type of comment, \( F(1, 86) = 4.97, p = .028, \eta_p^2 = .05 \), indicating that—in line with our predictions—participants inferred a lower dispositionality after reading ironic (\( M = 4.21, SD = 1.08 \)) compared to literal comments (\( M = 4.53, SD = 1.07 \)). We found no main effect of behavioral valence, \( F(1, 86) = 0.80, p = .37 \). We did, however, find an interaction effect between behavioral valence and type of comment, \( F(1, 86) = 4.46, p = .038, \eta_p^2 = .05 \). Pairwise comparisons with Bonferroni corrections showed that participants infer a significantly lower dispositionality from ironic compared to literal comments in the positive behavior condition, \( p < .01 \). In the negative behavior condition, however, we observed no differences between comment types, \( p = .95 \).

For generalizability, we observed main effects of both types of comment, \( F(1, 86) = 4.78, p = .031, \eta_p^2 = .05 \), and behavioral valence, \( F(1, 86) = 232.81, p < .001, \eta_p^2 = .73 \). In line with predictions, inferred generalizability was lower for ironic (\( M = 3.94, SD = 1.22 \)) than literal comments (\( M = 4.26, SD = 1.09 \)) and lower for negative (\( M = 3.05, SD = 1.21 \)) than for positive behavior (\( M = 5.15, SD = 1.03 \)). We observed no interaction effect between behavioral valence and type of comment, \( F(1, 86) = 0.002, p = .97 \).

**Discussion**

Taken together, the results of Experiment 4 show that the use of ironic versus literal comments has important communicative consequences in that they bias the inferences recipients draw about the target of the comments. Compared to literal statements, irony leads to more external attribution, as visible in lower inferred dispositionality and generalizability (H3). Irony also reduced inferred repetition likelihood compared to literal statements, but only in situations in which the comments refer to positive behavior. These results imply that from ironic comments about someone’s behavior, recipients infer that the behavior is atypical for the target, and not informative about stable dispositional characteristics of the target. These findings are in line with our arguments on the irony bias that irony activates the stereotypic expectancy in recipients, and makes them perceive the (stereotype-inconsistent) target situation as atypical. In such situations, the existing stereotypic category typically remains unchanged despite the presence of counter-stereotypic information (Bodenhausen et al., 1995).

These results are qualified by two important findings. First, results demonstrate that the differential inferences from ironic and literal comments are more pronounced in conditions in which they evaluate positive rather than negative behaviors. One reason for this finding can be that, in negative situations, irony is typically used as a politeness strategy to mitigate a negative evaluation (e.g., Brown & Levinson, 1987; Dews & Winner, 1995; Dews & Winner, 1999). If recipients infer that an ironic comment is
made for politeness reasons, it makes sense that they infer less information about the target or the speaker’s expectancies, because speakers may be hiding their true thoughts about the target.

Second, and apart from comment type, the repetition likelihood and generalizability inferences from positive behavior are higher than those from negative behavior. These results reflect a well-documented positivity bias (Klar & Giladi, 1997; Mezulis, Abramson, Hyde, & Hankin, 2004) that recipients believe that events usually turn out more positive than negative.

The results from Experiment 4 support our reasoning on the irony bias. We argue that the increased use of ironic comments about stereotype-inconsistent, as compared to stereotype-consistent behaviors (Experiments 1-3) functions as a means to implicitly communicate stereotype expectancies to message recipients. Experiment 4 demonstrated that ironic, compared to literal, comments indeed induce recipients to infer that the behavior is less likely caused by the actor’s stable dispositional characteristics, less likely to generalize to other situations, and, for positive behaviors, less likely to be repeated. An ironic comment about stereotype-inconsistent behavior can thus serve as a means to implicitly communicate that this behavior is atypical for the target, and consequently maintain the stereotypic expectancies about the target that instigated the ironic comment.

**General Discussion**

Social stereotypes are communicated through interpersonal communication, but little is known about the linguistic and communicative parameters through which this occurs (Harwood, 2010). This paper introduced the irony bias, and thereby reveals the role of verbal irony in the communication and maintenance of social stereotypes. Four experiments provided empirical evidence for the hypotheses that speakers find irony more appropriate in stereotype-inconsistent than in stereotype-consistent situations. Irony also has important communicative consequences, because it leads to more external attribution than literal language. As such, this paper is the first to integrate knowledge about irony (e.g., Attardo, 2000; Giora & Fein, 1999; Wilson & Sperber, 2004) and biased language use in social stereotyping (e.g., Beukeboom, 2014; Lee, 2007; Maass, 1999; Wigboldus & Douglas, 2007).

Experiments 1, 2, and 3 used converging methods to demonstrate that irony is found more appropriate in situations in which social stereotypes are violated than in situations in which social stereotypes are confirmed. In stereotype-inconsistent situations, irony enables speakers to implicitly communicate their prior stereotypic expectancy as well as the violation of that expectancy. Mediation analyses indeed showed that irony appropriateness (Experiment 2) and the choice for ironic over literal comments (Experiment 3) increased as a function of perceived unexpectedness of the target’s behavior.

Experiment 4 demonstrated that using ironic comments about actors showing stereotype-inconsistent behavior has important communicative consequences. Without receiving any information about the actor, recipients inferred from ironic—compared
to literal—comments that the actor was less likely to repeat the specific behavior in the future and that the behavior was less typical and stable for the actor. This implies that irony induces external attribution: Ironic comments implicitly communicate that the behavior is a one-time event that is unconnected to stable dispositional characteristic of the target. Because recipients are unlikely to change their beliefs when they perceive counter-stereotypical information as atypical (Bodenhausen et al., 1995), an ironic comment about stereotype-inconsistent behavior likely contributes to the maintenance of the stereotype expectancy even though the expectancy was violated.

Taken together, the four experiments reported in this paper provide an important contribution to the study of stereotypes in interpersonal communication. We have extended linguistic bias theory by showing that, next to language abstraction (Wigboldus et al., 2000), explanations (Sekaquaptewa et al., 2003), and negations (Beukeboom et al., 2010), verbal irony is another linguistic device through which stereotypes are implicitly communicated and maintained interpersonally. This is an important finding, because it contributes to our understanding of how stereotypic beliefs about social categories become shared in society. A biased use of irony contributes to social stereotyping and may thereby increase prejudice against minority groups (Geschke et al., 2010; von Hippel et al., 1997). Uncovering the linguistic mechanisms through which stereotypic expectancies are communicated is an important first step in trying to counter these adverse effects.

A speaker’s production of ironic comments can result from mere cognitive activation of stereotypic associations. According to literature on stereotyping, upon perceiving an actor, people immediately activate mental representations associated with the actor’s social category (Lepore & Brown, 1997; Wigboldus, Dijksterhuis, & van Knippenberg, 2003). The activation of stereotype-consistent thoughts makes it likely that they are uttered in an explicit comment, which is potentially ironic when the target’s behavior is stereotype inconsistent. This view is also reflected in several linguistic theories on irony, as several linguists argue that ironic comments typically “echo” or allude to expectations that are violated in the target situation (Kumon-Nakamura et al., 1995; Wilson & Sperber, 2004).

The communicative consequences of the irony bias could arise from the way in which recipients process irony. When processing ironic comments like “Gee, he is really smart,” both the literal (smart) and the intended meaning (stupid) are activated and remain active in working memory during and after processing (Akimoto et al., 2012; Giora & Fein, 1999; Giora et al., 1998). These processing effects of irony are comparable to negation processing, as related to the negation bias (Beukeboom et al., 2010). Negations like he is not smart are processed in a similar way: Recipients first process the utterance as an affirmation (smart) after which the negation operator (not) is added. As a consequence, negations activate associations with the negated concept, and thus make the opposite of the intended message meaning more accessible in recipients (Giora et al., 2007; Mayo et al., 2004). Thus, both the use of negations and irony in comments on stereotype-inconsistent behaviors activate a stereotype-consistent meaning in recipients, and likely also in the speaker. In other words, both negations (Beukeboom et al., 2010) and irony reflect speakers’ social stereotypes and serve to remind both speaker and recipients of the expected state of affairs.
The literature on irony and the literature on linguistic bias provide further corresponding ideas about potential functions of the irony bias. Our experimental situations involved interpersonal communication with high common ground between the sender and addressee as irony is typically used when speakers and recipients are friends and share common ground (cf. Kreuz & Link, 2002; Pexman & Zvaigzne, 2004). Thus, irony is more often used in situations in which the speaker and recipient share cultural and social assumptions like social stereotypes.

Interestingly, stereotype-confirming effects of linguistic biases are also mainly expected when the sender and recipient share common ground about the stereotypic expectancies about a target (Beukeboom, 2014; Fiedler et al., 2003). Moreover, speakers who exhibit in-group-serving linguistic biases in their communications about others are more appreciated as good group members than speakers whose communication deviates from such linguistic biases (Assilaméhou & Testé, 2013). Together, these findings corroborate the idea that the irony bias serves to maintain and share knowledge within social groups, which may facilitate relationships and group solidarity.

While many communication studies have shown how stereotypes (Givens & Monahan, 2005; Ramasubramanian, 2011; Schemer, 2012) and counter-stereotypes (Mastro & Tukachinsky, 2011; Ramasubramanian & Oliver, 2007) in the media serve to transmit and communicate stereotypic expectancies, the linguistic means underlying such stereotype communication have received less attention (Harwood, 2010). By using linguistic bias theory and expanding this theory with the irony bias, this paper highlights the important role of language in driving stereotype communication and maintenance.

The introduction of the irony bias opens up various opportunities for further research. First, our studies provide empirical evidence for the irony bias in an experimental setting. To further bolster external validity of our findings, our experimental approach could be supplemented by observational studies focusing on spontaneous irony usage in real-life situations. Furthermore, future research may investigate the role of irony in other contexts in which expectancies are violated. Language abstraction, for instance, has been shown to not only relate to stereotype expectancies (Wigboldus et al., 2000), but also to expectations regarding guilt attributions in criminal trials (Schmid & Fiedler, 1996), doctor-patient interactions (Watson & Gallois, 2002), and even perceptions of inanimate objects (Schellekens, Verlegh, & Smidts, 2013). Similar to language abstraction, a biased use of irony may thus serve similar effects across social contexts. Second, some ironic utterances may be easier to perceive as ironic than others. Experimental evidence suggests that ironic utterances with an explicitly evaluative word (e.g., an evaluative adjective) are perceived as easier than ironic utterances without such evaluative words (Burgers, van Mulken, & Schellens, 2012, Experiment 1). Furthermore, ironic utterances that are marked (with, for example, quotation marks, hyperbole, tag questions or with a different prosody) are also easier than ironic utterances without such markers (Burgers, van Mulken, & Schellens, 2012, Experiment 2; Bryant, 2010). Future research could thus combine such differential linguistic or prosodic choices in ironic utterances with our predictions of the irony bias.
Finally, in line with previous work on the linguistic expectancy bias (Wigboldus et al., 2000) and the negation bias (Beukeboom et al., 2010), we focused on the effects of stereotypic expectancies on irony use. Future research may, in line with work on the linguistic intergroup bias, also investigate how irony is used in (competitive) intergroup settings (e.g., ironic comments regarding targets from one’s in-group vs. out-group; cf. Maass, Salvi, Arcuri, & Semin, 1989).

The present study provides a sound basis for this type of future research that hopefully further extends our knowledge on the subtleties in language use that contribute to the maintenance of stereotypic representations in interpersonal communication. By revealing the mechanisms of linguistic biases, people may become more aware of biased language use, which in turn could prevent potentially negative effects.

Acknowledgments
The authors would like to thank Stefan Boerlage, Marlou van Boheemen, Stefanie Hamelink, Nikita Hendriks, and Berend Meijer for their assistance with Experiment 1 and Robin ten Hove for his invaluable assistance with stimulus design and data collection for Experiments 2 and 3.

Authors’ Note
A previous version of this paper was presented at the 63rd Annual Conference of the International Communication Association (ICA), London (UK), June 18, 2013.

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
The authors received no financial support for the research, authorship, and/or publication of this article.

Notes
1. We conducted a post-experimental test to check whether participants indeed perceived the ironic utterances as ironic and the literal utterances as literal. A total of 50 native speakers and first-year BA students of communication (M.age = 19.52, SD.age = 1.64, 82.0% female) completed an online questionnaire for course credit. Participants rated each comment on a 7-point semantic differential scale from 1 = completely literal to 7 = completely ironic. To make sure participants only judged the comments relative to the situation and apart from the stereotypic actors, we only presented them with the situation (e.g., successful installation/failure to install a new computer). Paired-sample t tests showed that, for each literal-ironic comment combination, the ironic utterances were perceived as more ironic than the literal utterances (with a certainty of at least p < .001). All ironic comments were rated highly above the scale midpoint (all Ms > 5.50), and all literal comments were rated below the scale midpoint (all Ms < 3.00).
2. Due to multicollinearity, we could not test for mediation of rated expectedness (Baron & Kenny, 1986). Stereotype consistency and the proposed mediator of expectedness were
strongly related ($\beta = -.64$), and collinearity diagnostics scores showed high variance proportions for both variables on the same eigenvalue (stereotype consistency: .69; expectedness: .88). Consequently, too little unique variance in the mediator variable remained to estimate a reliable mediation model.

3. Our analysis also provided a number of analyses that generalized over literal and ironic comments. We found no main effects of behavioral valence, $F(1, 100) = 1.01, p = .32$, or stereotype consistency, $F(1, 100) = 0.87, p = .35$. We did find a significant interaction between stereotype consistency and behavioral valence, $F(1, 100) = 5.61, p = .020, \eta^2_p = .05$. Pairwise comparisons with Bonferroni corrections showed no differences in the appropriateness of comments in stereotype-consistent and stereotype-inconsistent negative behavior ($p = .30$). For positive behavior, comments were considered more appropriate in stereotype-inconsistent ($M = 4.27$, $SD = 0.84$) than in stereotype-consistent situations ($M = 3.52$, $SD = 1.13, p < .05$).

4. In the same post-experimental test reported in Note 1, we also asked participants to rate the utterances used in Experiment 2. Again, paired-sample $t$ tests showed that, for each literal-ironic comment combination, the ironic utterances were perceived as more ironic than the literal utterances (with a certainty of at least $p < .001$). All ironic comments were rated highly above the scale midpoint (all $M$s > 5.50), and all literal comments were rated below the scale midpoint (all $M$s < 2.60).

5. Our analysis also provided a number of analyses that generalized over literal and ironic comments. We found no main effect of behavioral valence on the appropriateness of comments in general, $F(1, 57) = 1.37, p = .25$. Participants considered comments generally more appropriate in the stereotype-consistent ($M = 4.01$, $SD = 0.56$) than in the stereotype-inconsistent situations, $M = 3.87$, $SD = 0.54$, $F(1, 57) = 11.78, p = .001, \eta^2_p = .17$. The main effect of stereotype consistency was qualified by a significant interaction between stereotype consistency and behavioral valence, $F(1, 57) = 5.90, p = .018, \eta^2_p = .09$. Pairwise comparisons with Bonferroni corrections showed no differences in appropriateness of comments in stereotype-consistent and stereotype-inconsistent positive behavior ($p = .40$). For negative behavior, however, comments were considered more appropriate in stereotype-consistent ($M = 4.09$, $SD = 0.67$) than in stereotype-inconsistent situations ($M = 3.86$, $SD = 0.70, p < .01$).

6. Judd, Kenny, and McClelland (2001) also argue that if the sum scores are a significant predictor in this model, this implies a moderation effect. As the sum scores for both the models with the appropriateness of irony, $\beta = .004, t(55) = 0.03, p = .98$, and the appropriateness of literal statements, $\beta = .05, t(55) = 0.42, p = .68$, as dependent variables were non-significant, no moderation was found.

7. As the sum score was non-significant, $\beta = .13, t(47) = 0.91, p = .37$, no moderation was found.

**Supplemental Material**

The online appendices are available at http://crx.sagepub.com/supplemental

**References**


Klar, Y., & Giladi, E. E. (1997). No one in my group can be below the group’s average: A robust positivity bias in favor of anonymous peers. *Journal of Personality and Social Psychology, 73*, 887-901. doi:http://dx.doi.org/10.1037/0022-3514.73.5.885


**Author Biographies**

**Christian Burgers,** (PhD 2010, Radboud University Nijmegen, The Netherlands), is an assistant professor at the Department of Communication Science at VU University Amsterdam. In his research, he focuses on the use and effects of different types of figurative language, such as metaphor, hyperbole, and irony.

**Camiel J. Beukeboom,** (PhD 2003, VU University Amsterdam, The Netherlands) is an assistant professor at the Department of Communication Science at VU University Amsterdam. His research interests are linguistic biases and language use in interpersonal communication.