

The Negation Bias in Stereotype Maintenance: A Replication in Five Languages

Journal of Language and Social Psychology

1–18

© The Author(s) 2019



Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/0261927X19869759

journals.sagepub.com/home/jls

Camiel J. Beukeboom¹ , Christian Burgers¹,
Zsolt P. Szabó² , Slavica Cvejic³, Jan-Erik M.
Lönnqvist⁴ and Kasper Welbers¹

Abstract

Research on linguistic biases shows that stereotypic expectancies are implicitly reflected in language and thereby subtly communicated to message recipients. Research on the Negation Bias shows that the use of negations (e.g., *not stupid* vs. *smart*) is more pronounced in descriptions of stereotype-inconsistent compared with stereotype-consistent behaviors. This article reports a replication study of the original research conducted in Dutch, using newly developed materials, and in five different languages: English, Dutch, Hungarian, Finnish, and Serbian. The results validate the existence of the Negation Bias in all five languages. This suggests that negation use serves a similar stereotype-maintaining function across language families.

Keywords

intergroup communication, stereotypes, social category, negation, language, linguistic bias

Language use plays a crucial role in the formation and maintenance of social-category stereotypes (Beukeboom, 2014; Maass, 1999). Research has revealed a number of linguistic biases that show how speakers' stereotypic expectancies result in systematic

¹Vrije Universiteit Amsterdam, Amsterdam, Netherlands

²ELTE University, Budapest, Hungary

³University of Belgrade, Belgrade, Serbia

⁴University of Helsinki, Helsinki, Finland

Corresponding Author:

Camiel J. Beukeboom, Department of Communication Science, Vrije Universiteit Amsterdam, De Boelelaan 1081, 1081 HV Amsterdam, Netherlands.

E-mail: c.j.beukeboom@vu.nl

variations in language use when describing behaviors of categorized individuals. These biased descriptions, in turn, contribute to the consensualization of stereotypes within cultural groups (for a review, see Beukeboom & Burgers, 2019).

One linguistic bias is the Negation Bias (Beukeboom, Finkenauer, & Wigboldus, 2010). Research on the Negation Bias reveals that the use of negations (compared with affirmations) is more pronounced in descriptions of stereotype-inconsistent compared with stereotype-consistent behaviors. Negations (e.g., *The junkie was not deceitful*) imply exceptions to the rule and simultaneously introduce stereotype-consistent concepts in communications about stereotype-inconsistent information. Consequently, negation use—in behavior descriptions of categorized individuals—tends to reaffirm and maintain existing stereotypes.

In this article, we aim to find further evidence for the negation bias by conducting a conceptual replication study of Beukeboom et al. (2010, Study 2) using larger samples and newly developed materials. Moreover, we aim to extend the original work by testing whether the Negation Bias (originally obtained across multiple experiments in the Dutch language) can be validated in other languages. The present study was conducted in five different languages from three language families: English and Dutch (Germanic languages), Hungarian and Finnish (Finno-Ugric language family), and Serbian (Slavic language family). This allowed us to test whether the Negation Bias functions in a similar manner across different types of languages.

Linguistic Biases and the Negation Bias

To understand how stereotypes become shared knowledge, it is crucial to focus closely on language use in communications about socially categorized people (e.g., Beukeboom, 2014; Beukeboom & Burgers, 2019; Collins & Clément, 2012; Maass, 1999; Wigboldus & Douglas, 2007). In often quite subtle ways, language reflects, constructs, and maintains beliefs about social categories. With regard to describing behaviors of categorized individuals, Beukeboom and Burgers (2019) distinguished two types of stereotype-maintaining biases. One type relates to *what* information is communicated, and predicts that speakers tend to predominantly communicate about stereotype-consistent (rather than stereotype-inconsistent) information (e.g., Bratanova & Kashima, 2014). The second type of stereotype-maintaining biases, which we focus on here, relates to *how* information about categorized individuals is formulated. For example, the Linguistic Intergroup Bias (Maass, Salvi, Arcuri, & Semin, 1989), Linguistic Expectancy Bias (Wigboldus, Semin, & Spears, 2000), Stereotypic Explanatory Bias (Sekaquaptewa, Espinoza, Thompson, Vargas, & von Hippel, 2003), and Irony Bias (Burgers & Beukeboom, 2016) all show how subtle variations in verbal formulations communicate what is (un)expected for a categorized target or category.

The significance of these linguistic biases in descriptions of others lies in the fact that they implicitly communicate stereotypes to message recipients. The language used in stereotype-inconsistent messages (i.e., concrete, explanatory, ironic, negations) causes recipients to infer lower essentialism for the described behavior, which implies that the behavior is seen as unexpected and likelier caused by situational

circumstances than by enduring dispositional factors. In contrast, the language used in stereotype-consistent messages (i.e., abstract, unmarked, literal, affirmations) implies that the behavior is expected and likelier caused by the actor's enduring dispositional characteristics than by situational circumstances. Thus, through subtle variations in language use, speakers implicitly transmit their stereotypic expectancies to recipients, leading to the stereotype being shared and maintained interpersonally (Beukeboom, 2014; Beukeboom & Burgers, 2019; Maass, 1999; Wigboldus & Douglas, 2007).

The Negation Bias (Beukeboom et al., 2010) fits this line of work. It hypothesizes that when people describe stereotype-inconsistent behavior of an actor, they will be more likely to use negations than when they describe stereotype-consistent behavior (Beukeboom et al., 2010). The linguistics literature posits that negations' main function in natural language is denial, either of something explicitly stated or of an implicit pre-supposition implied in context (Tian & Breheny, 2016; Tottie, 1991). For example, a negation like "the train was not late this morning" is likelier used when the train is normally late than when it is normally punctual. Similarly, a statement like "Arsenal did not win" implies that someone expected that Arsenal would win, which is subsequently denied (Jordan, 1998; Moxey & Sanford, 2000). Thus, negations are meaningful in that they indicate that something is different, unusual, or contrary to an existing (implicit) expectancy (Jordan, 1998).

Experimental studies on processing and comprehension corroborate that negations are relatively more appropriate in communication about expectancy-inconsistent information. For instance, sentences with negations typically take longer to process and comprehend than affirmation sentences (Carpenter & Just, 1975; Deutsch, Gawronski, Strack, 2006; Gough, 1965), but this difference disappears when negations are used in an appropriate context (Glenberg, Robertson, Jansen, & Johnson-Glenberg, 1999; Tian & Breheny, 2016). One appropriate context (i.e., "a context of plausible denial"; Wason, 1965) is when an expectation needs to be noted or denied. Thus, negations seem more appropriate, and easier to process and comprehend, when denoting information that is inconsistent with expectations.

The Negation Bias (Beukeboom et al., 2010) links these ideas to stereotypes. When communicating about other people's behavior, people automatically activate social-category stereotypes associated with a discussed target, either primed by a target's appearance or by a category label used in the communication (Devine, 1989; Dijksterhuis & Van Knippenberg, 1996; Fiske, 1998; Lepore & Brown, 1997). Such (implicitly) activated social-category stereotypes facilitate the use of associated concepts in language use (e.g., junkies–deceitful). This can be in the form of an affirmation (e.g., the junkie is deceitful) to describe stereotype-consistent behavior, or a negation (e.g., the junkie was not deceitful), when target behavior is inconsistent with the activated stereotype. Moreover, following the pragmatic function of negations described above, in case of stereotype-inconsistent behavior, negations can function to denote an exception to this implicitly activated stereotype. Based on this, the original work on the Negation Bias (Beukeboom et al., 2010) hypothesized that negation use is likelier when a target person's behavior is more unexpected.

Beukeboom et al. (2010) provided evidence for the Negation Bias in three experiments showing that people have a higher preference for negations to describe stereotype-inconsistent (vs. stereotype-consistent) behaviors of others. The first study was a within-participants, forced-choice experiment. Participants were presented with a series of sentence pairs describing an actor's behavior. Each sentence in a pair described the same actor and behavior and had the same structure: one sentence contained a negated adjective (e.g., *The rock musician did not leave a messy hotel room*) and the other contained an affirmative antonym (e.g., *The rock musician left a neat hotel room*). Participants were asked to choose the sentence of each pair of which the choice of words felt most natural. Stereotype consistency was manipulated by varying the social category of the actors. The same sentence pairs were thus presented once with an actor for whom the described behavior was stereotypically unexpected (e.g., *rock musician—leaving a neat hotel room*), and once with an actor for whom the same described behavior was stereotypically expected (e.g., *cleaner—leaving a neat hotel room*). Results showed that participants were significantly more likely to choose the negation alternative in sentence pairs in which the actor's behavior was stereotype-inconsistent, compared with sentence pairs in which the actor's behavior was stereotype-consistent.

The second experiment used new stimulus sets (other actors and stereotypical behaviors) and the actor's behavior in the sentence, and descriptions of this behavior (either using a negation or affirmation) were separated. That is, participants were presented with a factual sentence about an actor's behavior (e.g., *The professor scored high on the IQ test*), followed by two possible descriptions: a negation description (e.g., *He is not stupid*) and an affirmation description (e.g., *He is smart*). For both descriptions, participants indicated on 7-point scales the extent to which they found it an appropriate description to use when describing the behavior in the sentence. Again, stereotype consistency was manipulated by varying the social category of the actor in the behavioral sentence (e.g., *professor, garbage man*). Results revealed a significant interaction between stereotype consistency and description type. In line with the Negation Bias hypothesis, participants rated negation descriptions (e.g., *He is not smart*) as more appropriate to describe stereotype-inconsistent behaviors (e.g., *The professor scored low on the IQ test*) than for stereotype-consistent behaviors (e.g., *The garbage man scored low on the IQ test*). In addition, for affirmation descriptions the opposite pattern was observed; affirmation descriptions (e.g., *He is smart*) were rated as more appropriate to describe stereotype-consistent behaviors (e.g., *The professor scored high on the IQ test*) than stereotype-inconsistent behaviors (e.g., *The garbage man scored high on the IQ test*).

The third experiment tested whether people spontaneously produce the Negation Bias when they are free to formulate their own person descriptions. Participants spontaneously described their impression of a male or female person performing either the stereotypically female sport jazz ballet, or the stereotypically male sport rugby. Results showed that participants spontaneously used more negations in descriptions of persons performing a stereotype-inconsistent behavior (*Man performing jazz ballet and woman*

playing rugby) than in descriptions of persons performing stereotype-consistent behavior (*Man playing rugby* or *woman performing jazz ballet*).

In addition to the effects of the experimental manipulations, mediational (Study 2) and correlational analyses (Study 3) demonstrated that perceived unexpectedness of the described behavior was positively related to (rated appropriateness of) negation usage. Finally, a fourth experiment focused on what message recipients infer from behavior descriptions containing either a negation or affirmation. This showed that, compared with affirmations, negations induce recipients to infer that the described characteristics were unexpected and of lower essentialism (i.e., less enduring, dispositional) for the target. Together, this provided evidence for the notion that negation use varies as a function of stereotypic expectancies, induces stereotype-confirming inferences in recipients, and thereby plays a role in stereotype transmission and maintenance.

In the present article, we aim to replicate and extend the main study on biased negation production (i.e., Beukeboom et al., 2010, Study 2). Many scholars nowadays agree that replication studies are essential for cumulative theoretical development (Brandt et al., 2014; Open Science Collaboration, 2015). Well-conducted replications, using large samples, can provide confirmation or disconfirmation of published results, and, when successful, greater confidence about the veracity of the predicted effect. This is also what we aim to achieve with regard to the Negation Bias.

Second, we extend the original study by testing whether the Negation Bias (originally obtained in Dutch) generalizes to other languages. Interestingly, all human languages have elements to express negation (Horn, 1989; Tian & Breheny, 2016). As opposed to affirmations, negation descriptions contain a negation marker (such as *not* or *no*) to deny the truth value of a particular proposition. Languages usually have more than one grammatical way to express negation. The Negation Bias focuses on syntactic negations that are expressed by means of a separate particle like NOT (e.g., *not smart*). We conducted five identical replication studies in five different languages belonging to different language families: English and Dutch (Germanic language family), Finnish and Hungarian (Finno-Ugric language family), and Serbian (Slavic language family). Each of these languages allow for syntactic negations, for example, *not friendly* is translated as, respectively, *niet vriendelijk* (Dutch), *ei ystävällinen* (Finnish), *nem barátságos* (Hungarian), and *nije druželjubiv* (Serbian). Because it can be expected that the pragmatic and cognitive mechanisms behind the Negation Bias function similarly across languages, we hypothesized that the Negation Bias would replicate in each of these languages.

Third, we extend the original study by testing whether the Negation Bias generalizes to other materials with other social categories and stereotypes. We developed new material sets with social-category stereotypes that were applicable across languages. This also allowed us to slightly improve the materials. While in the original study (Beukeboom et al., 2010) a few stimuli were based on general expectancies about situations (e.g., man attending a birthday party vs. funeral) rather than stereotypic expectancies (i.e., expectancies about social categories), the present new materials only included social categories, as this is the focus of the negation bias.

Method

We provide all materials and report how we determined sample size, all data exclusions (if any), all manipulations, and all measures in the studies.

Participants

Five participant samples of adult native speakers of English, Dutch, Hungarian, Finnish, and Serbian were recruited to complete an online Qualtrics questionnaire in their native languages. The survey link was spread via email to a Dutch panel *Kieskompas*,¹ and/or via social media (Hungarian, Finnish, Serbian). These participants were not paid, but were offered a ticket in a lottery for cinema vouchers. English participants were recruited in the United States from the Prolific panel (www.prolific.ac) and were paid as part of their panel membership.

Our design (see below) employs 6 material sets each with 2 observations in both the stereotype-inconsistent and stereotype-consistent conditions. We aimed for sample sizes > 100 (no maximum was set), so that we would have at least 1,200 ($6 \times 2 \times 100$) observations per condition. Based on the large effect obtained in the original study (Beukeboom et al., 2010, Study 2),² we estimated that this sample size enabled a sufficiently powered study (Brysbaert & Stevens, 2018). In all samples, analyses were only conducted after data collection had ended.³

Before analyzing data, we excluded participants who failed to meet our predetermined inclusion criteria: that is, age < 18 years, nonnative speakers of the study language, incomplete questionnaires, participants who took extremely long ($> a$ full day). Only the English version included attention-check items (Oppenheimer, Meyvis, & Davidenko, 2009), based on which four cases were excluded. In both the English and Serbian sample, we excluded one case for being extremely fast in combination with series of equal responses. No other data were excluded.

Size and demographics of the final samples were as follows:

Dutch: $N = 234$; 71% male ($n = 167$); Age $M = 56.9$, $SD = 15.1$, range (21-92)

English (United States): $N = 111$; 56% male ($n = 62$); Age $M = 28.4$, $SD = 9.1$, range (18-59)

Hungarian: $N = 115$; 38% male ($n = 44$); Age $M = 30.6$, $SD = 10.4$, range (19-68)

Finnish: $N = 111$; 17% male ($n = 19$); Age $M = 28.7$, $SD = 7.5$, range (20-58)

Serbian: $N = 122$; 40% male ($n = 49$); Age $M = 33.9$, $SD = 13.0$, range (18-67)

See also Appendix Table A1 (available online) for an overview.

Design and Materials

The study employed a 6 (Material set) \times 2 (stereotype consistent vs. stereotype inconsistent) \times 2 (negation vs. affirmation) within-participants design. The method is a

conceptual replication of Study 2 in Beukeboom et al. (2010) with (largely) new stimulus materials. The main dependent variable was the perceived appropriateness of behavior descriptions with negations and affirmations.

Participants were presented with 24 sentences from 6 material sets, each describing an actor's behavior that was either stereotype consistent or inconsistent. Each set consisted of two actors described with a social-category label (e.g., *student*; *housewife*) and two behaviors (e.g., *leaves the dishes in the sink for a week*; *washes the dishes immediately after dinner*). One behavior was stereotype consistent for Actor 1 but inconsistent for Actor 2, whereas, for the second (mirrored) behavior, this was reversed. Each set of actors and behaviors allowed for creating four sentences, two of which were stereotype consistent and two of which were stereotype inconsistent, by varying combinations of actor and behavior in the sentence. Each sentence, in turn, was associated with two relevant trait descriptions (one negation and one affirmation, e.g., *She is not tidy*; *She is messy* or *She is not messy*; *She is tidy*). Traits used in the negation and affirmation descriptions were antonyms, such that the two relevant descriptions of each sentence provided semantically comparable descriptions of the behavior. Note that in this design, actors, behaviors, and the judged descriptions were identical across stereotype-consistent and inconsistent conditions.

Material sets were selected after a pilot test on Hungarian ($N = 39$) and Dutch ($N = 28$) student samples testing 10 potential sets, partly based on the original Beukeboom et al. (2010) materials. Participants judged the actor–behavior sentences on unexpectedness and actor–trait associations using the same items as in the manipulation checks reported below. We selected sets in which—in both the Dutch and Hungarian samples—stereotype-inconsistent actor–behavior sentences were evaluated as significantly more unexpected than stereotype-consistent actor–behavior sentences, for both actors. Also, in all selected sets, actor–trait associations were higher for stereotype-consistent (vs. inconsistent) pairings. We assumed that the results of the Dutch and Hungarian pilot test would also apply to the other language samples; this was confirmed in the manipulation checks of the actual experiment. See Appendix Tables A2 and A3 for selected sets and manipulation check scores per item, for all five language samples.

All instructions and materials were translated from English into the other four languages by native speakers of each language (see <https://osf.io/tvre4/> for all materials and the appendix with additional data and analyses).

Procedure

The questionnaire first explained that the study focused on behavior descriptions, and that participants would be presented with sentences describing a person's behavior followed by two possible descriptions of that behavior. Participants were asked to indicate how appropriate they found each description to describe the behavior in the sentence. Subsequently, the sentences were presented (e.g., *The professor scores high on the IQ test*), each followed by two possible descriptions on the same page: a negation description (e.g., *He is not stupid*) and an affirmation description below (e.g., *He*

is smart). Participants indicated how appropriate they found each description on a 7-point scale ranging from 1 = *very inappropriate* to 7 = *very appropriate*. Sentences were presented in a mixed random order: a random sentence from Set 1 was followed by a random sentence from Set 2, and so forth till Set 6, and again starting with a remaining sentence from Set 1, until all 24 sentences were presented.

Next, we measured our manipulation check variable *expectedness of the actor's behavior* in the sentence. The behavior sentences were presented once more, with order randomized in the same manner as above, and participants indicated how expected they found the behavior for the person in the sentence on 7-point scales ranging from -3 = *very unexpected*, to +3 = *very expected* (recoded to 1-7 in results).

Next, we measured *actor-trait associations*. This variable is not included in the analyses, but results are reported in the Appendix, Table A3, as additional information about the stimuli. Participants were asked to rate the degree to which they associated the persons they just saw in the sentences with certain characteristics. This was asked for the two actors and two traits of each material set, again in mixed randomized order. For each of the four combinations per set, participants were asked to which degree they in general associated actors (e.g., *professors*) with the trait (e.g., *smart*), and answered on scales ranging from 1 = *not at all* to 7 = *very much*.

Finally, we asked participants to judge entitativity of each category⁴ and to report demographics.

Results

Manipulation Checks

In each language sample, we conducted a repeated-measures analysis of variance (ANOVA) on the mean rated expectedness of the stereotype-consistent and stereotype-inconsistent behavior sentences across the six material sets. We observed significant effects of stereotype consistency on mean-rated expectedness in

Dutch $F(1, 233) = 1799.25, p < .001, \eta^2_p = .89,$
 English $F(1, 110) = 600.53, p < .001, \eta^2_p = .85,$
 Hungarian $F(1, 114) = 626.43, p < .001, \eta^2_p = .85,$
 Finnish $F(1, 110) = 816.41, p < .001, \eta^2_p = .88,$ and
 Serbian $F(1, 121) = 515.07, p < .001, \eta^2_p = .81.$

In all language samples, participants rated the stereotype-consistent sentences as significantly more expected ($M_{\text{Dutch}} = 5.59, SD = 0.55; M_{\text{English}} = 5.56, SD = 0.58; M_{\text{Hungarian}} = 5.56, SD = 0.70; M_{\text{Finnish}} = 5.44, SD = 0.51; M_{\text{Serbian}} = 5.56, SD = 0.68$) than the stereotype-inconsistent sentences ($M_{\text{Dutch}} = 2.74, SD = 0.61; M_{\text{English}} = 2.99, SD = 0.69; M_{\text{Hungarian}} = 2.82, SD = 0.77, M_{\text{Finnish}} = 2.97, SD = 0.58; M_{\text{Serbian}} = 2.90, SD = 0.79$).

In all language samples, these differences were also significant ($p < .001$) for all six material sets separately. This confirms that our manipulation was successful. See

Appendix Tables A2 and A3 for detailed results per sentence and actor–trait associations in each set.

Appropriateness of Negation and Affirmation Behavior Descriptions

The Negation Bias predicts that the use of negations is perceived as relatively more appropriate in descriptions of stereotype-inconsistent (vs. stereotype-consistent) behaviors. To test this hypothesis, we computed the means of perceived appropriateness for the affirmation and negation descriptions for both the stereotype-inconsistent and consistent behaviors. This resulted in four scores which were included in a 2 (stereotype consistency of behavior: inconsistent vs. consistent) \times 2 (description type: affirmation vs. negation) ANOVA with repeated measures on both factors, which we employed in all language samples separately.

First, in line with the original study (Beukeboom et al., 2010, Study 2), in all language samples, we found a main effect of description type,

$$\begin{aligned} \text{Dutch } F(1, 233) &= 28.10, p < .001, \eta^2_p = .11, \\ \text{English } F(1, 110) &= 32.81, p < .001, \eta^2_p = .23, \\ \text{Hungarian } F(1, 114) &= 101.46, p < .001, \eta^2_p = .47, \\ \text{Finnish } F(1, 110) &= 37.35, p < .001, \eta^2_p = .25, \text{ and} \\ \text{Serbian } F(1, 121) &= 49.70, p < .001, \eta^2_p = .29. \end{aligned}$$

In all language samples, the use of affirmation descriptions was rated as significantly more appropriate ($M_{\text{Dutch}} = 4.86, SE = 0.06; M_{\text{English}} = 5.34, SE = 0.07; M_{\text{Hungarian}} = 4.89, SE = 0.08; M_{\text{Finnish}} = 5.05, SE = 0.08; M_{\text{Serbian}} = 4.90, SE = 0.08$) than the use of negation descriptions ($M_{\text{Dutch}} = 4.64, SE = 0.06; M_{\text{English}} = 4.92, SE = 0.08; M_{\text{Hungarian}} = 4.15, SE = 0.10; M_{\text{Finnish}} = 4.54, SE = 0.10; M_{\text{Serbian}} = 4.34, SE = 0.08$).

We also observed small significant main effects of stereotype consistency in Dutch, Hungarian, Finnish, and Serbian samples, but not in the English sample.

$$\begin{aligned} \text{Dutch } F(1, 233) &= 9.32, p = .003, \eta^2_p = .04, \\ \text{English } F(1, 110) &= 2.50, p = .12, \eta^2_p = .02, \\ \text{Hungarian } F(1, 114) &= 7.80, p = .006, \eta^2_p = .06, \\ \text{Finnish } F(1, 110) &= 9.25, p = .003, \eta^2_p = .08, \text{ and} \\ \text{Serbian } F(1, 121) &= 3.97, p = .05, \eta^2_p = .03. \end{aligned}$$

This peripheral finding indicates that descriptions of stereotype-inconsistent behavior sentences were judged as somewhat more appropriate ($M_{\text{Dutch}} = 4.80, SE = 0.06; M_{\text{English}} = 5.16, SE = 0.07; M_{\text{Hungarian}} = 4.57, SE = 0.09; M_{\text{Finnish}} = 4.84, SE = 0.08; M_{\text{Serbian}} = 4.67, SE = 0.08$) than descriptions of stereotype-consistent behaviors ($M_{\text{Dutch}} = 4.71, SE = 0.06; M_{\text{English}} = 5.10, SE = 0.07; M_{\text{Hungarian}} = 4.47, SE = 0.08; M_{\text{Finnish}} = 4.75, SE = 0.08; M_{\text{Serbian}} = 4.57, SE = 0.07$). Note that this finding stands in contrast to the stereotype consistency bias (Bratanova & Kashima, 2014) discussed above. It might be the case that stereotype-inconsistent (compared with consistent)

Table 1. Means (and Standard Deviations) of Judged Appropriateness of Negation and Affirmation Descriptions as a Function of Stereotype Consistency (Consistent, Inconsistent) of Described Behavior, in Dutch, English, Hungarian, Finnish, and Serbian language samples.

Appropriateness of behavior descriptions	Description type			
	Negations		Affirmations	
	Consistent behaviors	Inconsistent behaviors	Consistent behaviors	Inconsistent behaviors
In Dutch ($N = 234$)	4.58 ^a (0.96)	4.71 ^b (0.93)	4.84 ^x (0.88)	4.88 ^x (0.94)
In English ($N = 111$)	4.86 ^a (0.90)	4.98 ^b (0.90)	5.34 ^x (0.70)	5.34 ^x (0.73)
In Hungarian ($N = 115$)	3.94 ^a (1.07)	4.37 ^b (1.08)	4.99 ^x (0.94)	4.78 ^y (0.93)
In Finnish ($N = 111$)	4.47 ^a (1.06)	4.61 ^b (1.07)	5.03 ^x (0.85)	5.07 ^x (0.81)
In Serbian ($N = 122$)	4.15 ^a (0.94)	4.53 ^b (1.05)	4.99 ^x (0.89)	4.80 ^y (0.90)

Note. Means in rows with different superscript within negations (^{a, b}) and affirmations (^{x, y}) description types are significantly different according to Bonferroni-adjusted pairwise comparisons with a certainty of $p < .02$ in English data, and $p < .005$ in all other languages.

information is perceived as more appropriate because this information is new/unexpected and therefore more informative, in line with Grice's (1975) conversational maxims.

More relevant for the Negation Bias, however, is the interaction between stereotype consistency and description type. This interaction was significant for

Dutch $F(1, 233) = 4.35, p = .04, \eta^2_p = .02$,
 English $F(1, 110) = 6.14, p = .01, \eta^2_p = .05$,
 Hungarian, $F(1, 114) = 50.24, p < .001, \eta^2_p = .31$, and
 Serbian, $F(1, 121) = 28.70, p < .001, \eta^2_p = .19$,

and marginally significant for

Finnish, $F(1, 110) = 3.61, p = .06, \eta^2_p = .03$.

In line with the Negation Bias, pairwise comparisons with Bonferroni corrections (see Table 1) showed that, in all language samples, the use of negations was perceived as significantly more appropriate in descriptions of stereotype-inconsistent (vs. stereotype-consistent) behaviors. The affirmation descriptions only differed significantly in the Hungarian and Serbian samples, where the use of affirmations was judged as more appropriate in descriptions of stereotype-consistent compared to inconsistent behaviors, in line with the original study.

Alternative Analyses

The hypothesis tests mentioned above follow the methods for statistical testing in the original article (Beukeboom et al., 2010, Study 2). This test generalized across

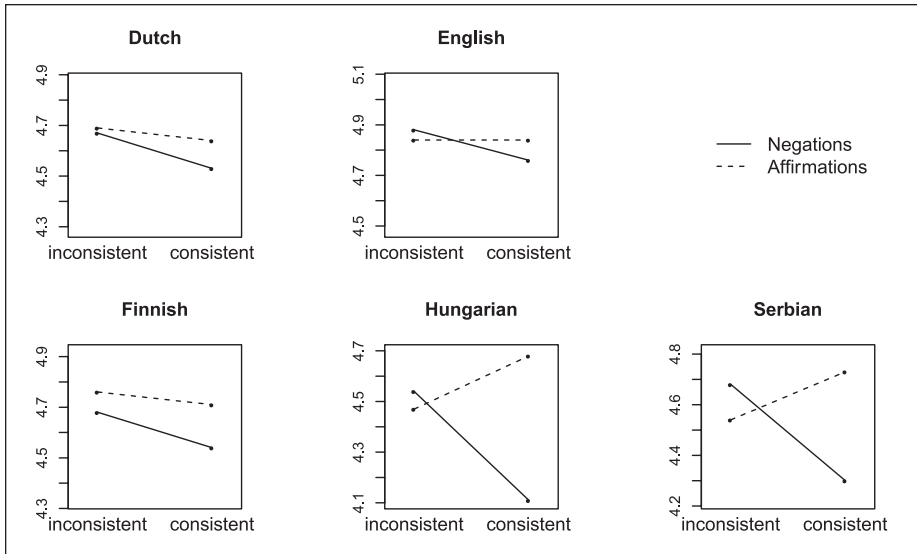


Figure 1. Fixed effects of stereotype consistency (inconsistent vs. consistent) on the judged appropriateness (y-axis) of negation and affirmation descriptions in the five language samples (full models in Appendix Tables A5-A9).

material sets and participants. In linguistics, scholars recommend to analyze similar designs with mixed-effects modelling with random crossed effects for stimulus sets and participants (Baayen, Davidson, & Bates, 2008). This analysis is advantageous, because it not only distinguishes fixed effects related to the independent variables, but also takes random effects due to (random) variation between different participants and different stimulus sets into account. To apply this procedure, we used R (R Core Team, 2017), with R packages *lme4* (1.1.21) for using linear mixed effects models (Bates, Maechler, Bolker, & Walker, 2015) and *sjPlot* (2.6.3) for interpreting the results (Lüdtke, 2017). The *sjPlot* package includes several established techniques for computing *p* values and model fit. For *p* values we used the conditional *F* tests with Kenward–Roger approximation, and the model fit is reported as the marginal and conditional *R*², using the method proposed by Nakagawa, Johnson, and Schielzeth (2017).

We first transposed the datasets of the five language samples such that each actor–behavior sentence became a case, leading to 24 cases per participant (6 stimulus sets × 4 sentences). Next, we estimated mixed-effects models with stereotype consistency as fixed factor, and random intercepts for participants and stimulus sets thus correcting for potential differences between stimulus sets and individual participants. See Appendix Tables A5-A9, first and third columns (black font) and Figure 1 for plots for each language sample. Results of this more precise analysis confirm the previously presented findings. We find a negative direct effect of stereotype consistency on the appropriateness of negation use in all five languages, indicating that negations are more appropriate to describe stereotype-inconsistent (vs. consistent) behaviors. This

confirms the Negation Bias and shows the findings are not due to the type of analyses conducted.

With regard to affirmations, we find a direct effect of stereotype consistency on the appropriateness of affirmation use for two languages (Hungarian, Serbian), indicating that affirmations are judged as more appropriate in description of stereotype-consistent (vs. inconsistent) situations. For the other three languages, we found no direct effect of consistency on the appropriateness of affirmations.

The same analyses also allowed us to further explore one hypothesized explanatory variable of the effect of stereotype consistency on appropriateness of negation (and affirmation) use; perceived expectedness of the actor's behavior. These exploratory analyses (see Appendix for a detailed description) provide evidence that perceived expectedness of the actor's behavior is an explanatory variable in predicting the appropriateness of negation (and affirmation) descriptions in Hungarian and Serbian, but not in the other language samples. This suggests there are other explanatory variables which will be discussed in the next section.

General Discussion and Conclusion

The findings corroborate the Negation Bias in five different languages. In all language samples, negation use was rated as more appropriate in descriptions of stereotype-inconsistent (vs. stereotype-consistent) behavior. Specifically, to describe the same behaviors (e.g., *washes the dishes immediately after dinner*), the use of negation descriptions (e.g., *not messy*) was judged as more appropriate when the actor was stereotypically unexpected to show this behavior (e.g., *student*) compared with when the actor was expected to show this behavior (e.g., *housewife*). These findings were confirmed both by the original statistical tests using repeated measures ANOVA (Beukeboom et al., 2010, Study 2) and advantageous mixed effects models correcting for random effects between participants and stimulus sets. This provides confirming evidence, using new experimental materials, for the notion that negations function to communicate that an actor's behavior is stereotype inconsistent. The fact that these findings replicate across two Germanic languages (Dutch, English), two Uralic languages (Finnish, Hungarian) and one Slavic language (Serbian) suggests that the Negation Bias serves a similar function in the transmission and maintenance of social stereotypes within these language families.

A number of findings are noteworthy in comparison with the original study.

First, the effect sizes were smaller than in the original study and also varied considerably between the five language samples. The effect size of the stereotype consistency \times description type interaction on description appropriateness was large in the original study ($\eta_p^2 = .46$), compared with the present Hungarian ($\eta_p^2 = .31$), Serbian ($\eta_p^2 = .19$), English ($\eta_p^2 = .05$), Finnish ($\eta_p^2 = .03$), and Dutch ($\eta_p^2 = .02$) samples. An important reason is that the interaction in the original study is for a large part driven by differences in affirmations (i.e., more appropriate for stereotype-consistent vs. inconsistent). Note that affirmations are not part of the Negation-Bias hypothesis which focuses on the different use of negations, but are included in the design to allow

comparing negation appropriateness in stereotype consistent and inconsistent situations while keeping actors, behaviors, and the judged descriptions the same across conditions. Nevertheless, the judged appropriateness of affirmations is interesting in itself. In the present study, we only observed significant effects for affirmations in the Hungarian and Serbian sample, and in all samples the difference is more pronounced for negations (See Tables 1, A5-A9 and Figure 1 for a visualization).

These differences may be due to the fact that the present studies used different materials than the original. Overall (across six sets) the effect sizes of the manipulation check (i.e., judgements of expectedness of behaviors in sentences) were large and comparable to the original study, but there obviously were differences in the strength of stereotypic expectancies between material sets within the different language samples. We aimed to manipulate more or less universal stereotypes in our materials, but such differences between cultural groups are unavoidable. Still, the fact that the findings were replicated in all language samples and were confirmed in our alternative analyses using a mixed effects model which corrected for potential differences between material sets and individual participants, substantiates the robustness of the negation bias effect.

Another difference is that, in the present study, data were collected in online experimental surveys rather than laboratory environments as in the original study. Although sample sizes in the present studies are much larger than in the original study, it is very well possible that a part of these online participants was relatively less focused on the questionnaire. The noise this introduces may also explain the lower effect sizes.

Finally, it is interesting to consider the underlying process of the Negation Bias effect. The increased use of negations to describe stereotype-inconsistent events may be driven by two explanatory variables that could function similarly across languages. The first relates to work in the field of pragmatics that argues that negations function to indicate something that is different, unusual, or contrary to an existing (implicit) presupposition or expectancy that is relevant in context (e.g., Jordan, 1998; Moxey & Sanford, 2000; Wason, 1965). In communication about categorized individuals, negations then function to denote an exception to implicitly activated stereotypes. This means that the use of negations should be likelier to the extent that a target person's behavior is more unexpected. This explanation does not pertain to the use of affirmations. In an additional exploratory analysis (see Appendix), we checked for the relation between expectedness of the described behavior and appropriateness of negation use and found mixed results (Appendix Tables A5-A9, second and fourth columns; gray font). As expected, and in line with the original study, for Hungarian and Serbian expectedness negatively predicted negation appropriateness. For Dutch, English, and Finnish, expectedness was unlike the (different type of) additional analyses in the original study, not related to negation appropriateness. This suggests that other explanatory variables besides expectedness play a role in driving the Negation Bias.

One second probable explanatory variable relates to work within the social-cognition field showing that social categorization automatically activates associated stereotypes (Devine, 1989; Dijksterhuis & Van Knippenberg, 1996; Fiske, 1998; Lepore & Brown, 1997). A categorization, for instance by means of a label, should cognitively activate stereotype consistent terms, and inhibit inconsistent terms, which makes their

use in person descriptions likelier. This means that the more strongly an activated category is associated with a given concept (e.g., *junkies–deceitful*), the likelier it will be used when describing a categorized target. This should both result in an increased use of negations in stereotype inconsistent situations and an increased use of affirmations in stereotype consistent situations. To test for the effect of this explanatory variable a measure of actor–trait associations is needed. We did measure actor–trait associations in our material sets, and these associations were in general very strong (see Table A3). Yet the nature of the current design (only 2 actor–traits scores per set and 4 actor–behavior sentences) prevents a proper mediation analysis using this variable (see Appendix).

All in all, the evidence for the underlying mechanism of the Negation Bias (both in the original and the present studies) is not conclusive. It seems likely that the biased use of negations may be driven by a combination of the above mechanisms. Future research might be able to further explore this and also test whether actor–trait associations (the social cognition explanation) can explain the differences in effects on affirmations we observed in the different language samples.

Implications and Future Research

The increased use of negations to describe stereotype-inconsistent behaviors has several implications for stereotype maintenance. First, in line with their pragmatic function, negations imply to message recipients that the described target's behavior is unexpected, an exception to the rule, and likelier caused by situational circumstances than by enduring dispositional factors (Beukeboom et al. 2010, Study 4). By describing information using a negation, people also convey a mitigated, more neutral version of the described event (Giora, Fein, Ganzi, & Alkeslassy Levi, 2005). For instance, using *not smart* conveys a less negative impression than stating *stupid* because the former introduces a positive concept. Likewise, *not stupid* conveys a less positive impression than *smart* by introducing a negative concept. Moreover, negations introduce stereotype-consistent concepts in communication about stereotype-inconsistent information, and thereby communicate what was expected instead. Thus, by describing stereotype-inconsistent behavior with negations rather than affirmations, the behavior is framed as unexpected, transient, more neutral in valence, while re-affirming existing associations with a category. The Negation Bias thereby functions to share and maintain social-category stereotypes (Beukeboom & Burgers, 2019). Interestingly, the present study suggests that this mechanism generalizes across various languages from different language families. Future studies might explore whether the Negation Bias also extends to non-European languages like Chinese, Arabic, and Hindi.

A number of boundary conditions are important. First, it should be noted that, although all languages have elements to express negation (Horn, 1989; Tian & Breheny, 2016), various grammatical forms to express negations exist (Verhagen, 2005), and these forms may differ between languages. The Negation Bias focuses on syntactic negations, which are expressed by means of a separate particle NOT (e.g., *not smart*) and can be found in most languages. It may work differently, however, for morphological negations, which are expressed by prefix (e.g., un- as in unhappy) or

suffix (e.g., *-less* as in *hopeless*). It has been argued that syntactic negations (“The clown is not happy”) are processed in two steps, with recipients first activating the negated concept (happy) followed by the negation (not), thereby activating the stereotypic inference that clowns are usually happy (e.g., Fraenkel & Schul, 2008; Giora et al, 2005; Mayo et al., 2004). In contrast, morphological negations (e.g., “The clown is unhappy”) are typically processed in one step, with addressees immediately coming to the intended meaning of “unhappy” without first activating the negated concept (“happy”; e.g., Mayo, Schul, & Burnstein, 2004). This could imply that the stereotype maintaining effect of the Negation Bias would primarily function for syntactic (but not for morphological) negations. Yet this may be different for different languages. Future research might take these different types of formulations into account and also further address recipient inferences (e.g., Beukeboom et al., 2010, Study 4) in addition to negation versus affirmation production.

Along with other linguistic variations, negations are an important linguistic means by which stereotypes are interpersonally shared and maintained. The present studies show that the Negation Bias functions similarly in various languages, just like it has for instance been shown for the Linguistic Intergroup Bias (Maass, 1999). Research into these processes is important, because awareness of the subtle linguistic ways through which people share potentially harmful stereotypes is needed for any attempt to prevent or correct them.

Acknowledgments

The authors would like to thank various anonymous reviewers for their valuable feedback on earlier drafts of this article.

Declaration of Conflicting Interests


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

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iDs

Camiel J. Beukeboom  <https://orcid.org/0000-0002-0364-2784>

Zsolt P. Szabó  <https://orcid.org/0000-0001-8124-2869>

Supplemental Material

Supplemental material for this article is available online.

Notes

1. *Kieskompas* (www.kieskompas.nl) is a Dutch Voting Advice Application (VAA), which can be used to get a voting advice for a relevant local, national, or European election.

- Users of this VAA had the opportunity to sign up for voluntary participation in academic research. A random sample of this panel was invited for participation via email.
2. The original study reports a stereotype consistency \times description type interaction effect with $\eta^2_p = .46$, which according to Richardson (2011) corresponds to a large effect (i.e., $\eta^2_p > .1379$) following Cohen's benchmarks.
 3. The English version was tested earlier twice, on a U.K. and a U.S. sample and did not confirm hypotheses. We think this should be attributed to a mistranslation in English (e.g., in both instructions and answering scale the term *applicable* rather than *appropriate* was used), which presumably lead to confusion about the task among participants. In the method presented here this was corrected.
 4. Category entitativity was not measured in the original study and included in relation to a different project; results are not reported here.

References

- Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*, 390-412. doi:10.1016/j.jml.2007.12.005
- Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*, 1-48. doi:10.18637/jss.v067.i01
- Beukeboom, C. J. (2014). Mechanisms of linguistic bias: How words reflect and maintain stereotypic expectancies. In J. Laszlo, J. P. Forgas, & O. Vincze (Eds.), *Social cognition and communication* (pp. 313-330). New York, NY: Psychology Press.
- Beukeboom, C. J., & Burgers, C. (2019). How stereotypes are shared through language: A review and introduction of the social categories and stereotypes communication (SCSC) framework. *Review of Communication Research*, *7*, 1-37. doi:10.12840/issn.2255-4165.017
- Beukeboom, C. J., Finkenauer, C., & Wigboldus, D. H. J. (2010). The negation bias: When negations signal stereotypic expectancies. *Journal of Personality and Social Psychology*, *99*, 978-992. doi:10.1037/a0020861
- Brandt, M. J., Iljerman, H., Dijksterhuis, A., Farach, F. J., Geller, J., Giner-Sorolla, R., . . . van 't Veer, A. (2014). The replication recipe: What makes for a convincing replication? *Journal of Experimental Social Psychology*, *50*, 217-224. doi:10.1016/j.jesp.2013.10.005
- Bratanova, B., & Kashima, Y. (2014). The "saying is repeating" effect: Dyadic communication can generate cultural stereotypes. *Journal of Social Psychology*, *154*, 155-174. doi:10.1080/00224545.2013.874326
- Brysbaert, M., & Stevens, M. (2018). Power analysis and effect size in mixed effects models: A tutorial. *Journal of Cognition*, *1*, 1-20. doi:10.5334/joc.10
- Burgers, C., & Beukeboom, C. J. (2016). Stereotype transmission and maintenance through interpersonal communication: The irony bias. *Communication Research*, *43*, 414-441. doi:10.1177/0093650214534975
- Carpenter, P., & Just, M. (1975). Sentence comprehension: A psycholinguistic processing model of verification. *Psychological Review*, *82*, 45-73. doi:10.1037/h0076248
- Collins, K. A., & Clément, R. (2012). Language and prejudice: Direct and moderated effects. *Journal of Language and Social Psychology*, *31*, 376-396. doi:10.1177/0261927X124446611
- Deutsch, R., Gawronski, B., & Strack, F. (2006). At the boundaries of automaticity: Negation as reflective operation. *Journal of Personality and Social Psychology*, *91*, 385-405. doi:10.1037/0022-3514.91.3.385
- Devine, P. G. (1989). Stereotypes and prejudice: Their automatic and controlled components. *Journal of Personality and Social Psychology*, *56*, 5-18. doi:10.1037/0022-3514.56.1.5

- Dijksterhuis, A., & Van Knippenberg, A. (1996). The knife that cuts both ways: Facilitated and inhibited access to traits as a result of stereotype activation. *Journal of Experimental Social Psychology*, *32*, 271-288. doi:10.1006/jesp.1996.0013
- Fiske, S. T. (1998). Stereotyping, prejudice, and discrimination. In D. T. Gilbert, S. T. Fiske & G. Lindzey (Eds.), *The handbook of social psychology* (4th ed, pp. 357-411). New York, NY: McGraw-Hill.
- Fraenkel, T., & Schul, Y. (2008). The meaning of negated adjectives. *Intercultural Pragmatics*, *5*, 517-540. doi:10.1515/IPRG.2008.025
- Giora, R., Fein, O., Ganzi, J., & Alkeslassy Levi, N. (2005). On negation as mitigation: The case of negative irony. *Discourse Processes*, *39*, 81-100. doi:10.1207/s15326950dp3901_3
- Glenberg, A. M., Robertson, D. A., Jansen, J. L., & Johnson-Glenberg, M. C. (1999). Not propositions. *Cognitive Systems Research*, *1*, 19-33. doi:10.1016/S1389-0417(99)00004-2
- Grice, H. P. (1975). Logic and conversation. In P. Cole & J. Morgan (Eds.), *Studies in syntax and semantics III: Speech acts* (pp. 183-198). New York, NY: Academic Press.
- Gough, P. B. (1965). Grammatical transformations and speed of understanding. *Journal of Verbal Learning and Verbal Behavior*, *4*, 107-111. doi:10.1016/S0022-5371(65)80093-7
- Horn, L. R. (1989). *A natural history of negation*. Chicago, IL: University of Chicago Press.
- Jordan, M. P. (1998). The power of negation in English: text, context, and relevance. *Journal of Pragmatics*, *29*, 705-752. doi:10.1016/S0378-2166(97)00086-6
- Lepore, L., & Brown, R. (1997). Category and stereotype activation: Is prejudice inevitable? *Journal of Personality and Social Psychology*, *72*, 275-287. doi:10.1037//0022-3514.72.2.275
- Lüdecke, D. (2017). sjPlot: Data visualization for statistics in social science. *R package version 2.4.0*. Retrieved from <https://CRAN.R-project.org/package=sjPlot>
- Maass, A. (1999). Linguistic intergroup bias: Stereotype perpetuation through language. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 31, pp. 79-121). San Diego, CA: Academic Press.
- Maass, A., Salvi, D., Arcuri, L., & Semin, G. R. (1989). Language use in intergroup contexts: The linguistic intergroup bias. *Journal of Personality and Social Psychology*, *57*, 981-993. doi:10.1037/0022-3514.57.6.981
- Mayo, R., Schul, Y., & Burnstein, E. (2004). "I am not guilty" vs. "I am innocent": Successful negation may depend on the schema used for its encoding. *Journal of Experimental Social Psychology*, *40*, 433-449. doi:10.1016/j.jesp.2003.07.008
- Moxey, L. M., & Sanford, A. J. (2000). Communicating quantities: A review of psycholinguistic evidence of how expressions determine perspectives. *Applied Cognitive Psychology*, *14*, 237-255. doi:10.1002/(SICI)1099-0720(200005/06)14:3<237::AID-ACP641>3.0.CO;2-R
- Nakagawa, S., Johnson, P. C., & Schielzeth, H. (2017). The coefficient of determination R^2 and intra-class correlation coefficient from generalized linear mixed-effects models revisited and expanded. *Journal of the Royal Society Interface*, *14*, 20170213. doi:10.1098/rsif.2017.0213
- Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, *349*, aac4716. doi:10.1126/science.aac4716
- Oppenheimer, D. M., Meyvis, T., & Davidenko, N. (2009). Instructional manipulation checks: Detecting satisficing to increase statistical power. *Journal of Experimental Social Psychology*, *45*, 867-872. doi:10.1016/j.jesp.2009.03.009
- R Core Team. (2017). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from <https://www.R-project.org/>

- Richardson, J. T. E. (2011). Eta squared and partial eta squared as measures of effect size in educational research. *Educational Research Review*, 6(2), 135-147. doi:10.1016/j.edurev.2010.12.00
- Sekaquaptewa, D., Espinoza, P., Thompson, M., Vargas, P., & von Hippel, W. (2003). Stereotypic explanatory bias: Implicit stereotyping as a predictor of discrimination. *Journal of Experimental Social Psychology*, 39, 75-82. doi:10.1016/S0022-1031(02)00512-7
- Tian, Y., & Breheny, R. (2016). Dynamic pragmatic view of negation processing. In P. Larrivée & C. Lee (Eds.), *Negation and polarity: Experimental perspectives* (pp. 21-43). Cham, Switzerland: Springer International. doi:10.1007/978-3-319-17464-8_2
- Tottie, G. (1991). *Negation in English speech and writing: A study in variation*. San Diego, CA: Academic Press.
- Verhagen, A. (2005). *Constructions of intersubjectivity: Discourse, syntax, and cognition*. Oxford, England: Oxford University Press.
- Wason, P. C. (1965). The contexts of plausible denial. *Journal of Verbal Learning and Verbal Behavior*, 4, 7-11. doi:10.1016/S0022-5371(65)80060-3
- Wigboldus, D., & Douglas, K. (2007). Language, stereotypes, and intergroup relations. In K. Fiedler (Ed.), *Social communication* (pp. 79-106). New York, NY: Psychology Press.
- Wigboldus, D. H., Semin, G. R., & Spears, R. (2000). How do we communicate stereotypes? Linguistic bases and inferential consequences. *Journal of Personality and Social Psychology*, 78, 5-18. doi:10.1037//0022-3514.78.1.5

Author Biographies

Camiel J. Beukeboom is an assistant professor in the Department of Communication Science at Vrije Universiteit Amsterdam. His research, broadly, focuses on interpersonal communication and language use, and in particular, on the role of language in the communication of social category perceptions and stereotypes.

Christian Burgers is an associate professor in the Department of Communication Science at Vrije Universiteit Amsterdam and a Professor by Special Appointment in Strategic Communication (Logeion Chair) at the University of Amsterdam (ASCoR). He studies strategic communication through indirect language (e.g., metaphor, irony, negation) across discourse domains.

Zsolt P. Szabó got involved in this project as a visiting scholar at VU Amsterdam. He holds a PhD in psychology from the University of Pécs and currently works as a lecturer at ELTE University Budapest.

Slavica Cvejić got involved in this project when she was visiting VU Amsterdam as a student. She holds a MA in English Language, Literature and Culture from the University of Belgrade and currently works as an English language teacher.

Jan-Erik M. Lönnqvist is a professor at the University of Helsinki. His research focuses on personality traits and values, and moral psychology.

Kasper Welbers holds a PhD in communication science from the Vrije Universiteit Amsterdam, where he is currently employed as a postdoctoral researcher. His research focuses on how people find and share news in the digital age, which he studies by tracking the diffusion of news messages using computational methods.