Linguistic Bias

Camiel J. Beukeboom, Department of Communication Science, VU University Amsterdam and Christian Burgers, Department of Communication Science, VU University Amsterdam

https://doi.org/10.1093/acrefore/9780190228613.013.439

Published online: 27 July 2017

Summary
Social categorization and stereotypes play a pervasive and fundamental role in social perception, judgment, and interaction. Although stereotypes are functional by allowing us to make sense of our complex social environment, their use can promote prejudice and discrimination when individuals are treated based on generic stereotypic expectancies, rather than on available individuating information. Prejudice and discrimination emerge from generalized (negative) stereotypic associations that people hold about social categories. These stereotypes become socially shared within (sub)cultures through communications about categorized people and their behavior. Research on biased language use reveals the communicative and linguistic processes through which stereotypes are formed and maintained. When communicating about other people and their behavior, our language echoes the existing stereotypic expectancies we have with categorized individuals (often without our conscious awareness). A linguistic bias is defined as a systematic asymmetry in word choice that reflects the social-category cognitions that are applied to the described group or individual(s). Three types of biases are distinguished in the literature that reveal, and thereby maintain, social-category cognitions and stereotypes.

First, when labeling individuals, the types of category labels that we choose reflect existing social category cognitions. Second, once target individuals are labeled as members of a category, people tend to communicate predominantly stereotype-congruent information (rather than incongruent information). Third, research has revealed several biases in how we formulate information about categorized individuals. These formulation differences (e.g., in language abstraction, explanations, use of negations, irony) subtly reveal whether a target’s behavior was stereotypically expected or not. Behavioral information that is in line with social-category knowledge (i.e., stereotype-consistent) is formulated differently compared to incongruent information (i.e., stereotype-inconsistent).

In addition, when communicating with individuals we have categorized in a given social category, our language may subtly reveal the stereotypic expectancies we have about our conversation partners. It is important to be aware of these stereotype maintaining biases as they play an important role in consensualizing both benevolent and harmful stereotypes about social categories.

Keywords: social categories, stereotypes, language, linguistic bias, linguistic intergroup bias, negation bias, irony bias, prejudice, discrimination, intergroup communication

Subjects: Intergroup Communication
Social Category Stereotypes and Language Use

Grouping individuals into categories is a fundamental—and functional—human tendency. Social categories, and the stereotypes (i.e., the knowledge and expectancies about probable behaviors, features, and traits) we associate with them, help people to make sense of their social world and to gain some predictability (Mackie, Hamilton, Susskind, & Rosselli, 1996; Moskowitz, 2005). For instance, when classifying somebody as a nurse, some types of behavior, characteristics, and traits (e.g., typically female, white coat, cares for ill people, friendly) may be expected, while other types of behaviors and traits (e.g., puts out fires, male, aggressive) are unexpected.

After we have identified an individual as a member of a given social category, we can thus draw on knowledge that we have come to associate with that category. The simplification this brings, however, also has serious downsides. When relying on social category knowledge, people tend to exaggerate similarities between individuals within categories (i.e., they are seen as all alike) as well as differences between categories. Moreover, when applying generic stereotypic expectancies to categorized persons, people tend to ignore the individuality of and situational constraints on the person (Allport, 1954; Mackie et al., 1996). Prejudice and discrimination generally refer to situations in which individuals are treated, described, and/or pre-judged based on such generic stereotypic expectancies and sentiments, rather than on available individuating information.

The negative consequences of relying on social-category stereotypes are even more substantial because stereotypes become consensually shared across large numbers of people. Within the (sub)cultures in which we reside, we learn which social categories are considered meaningful, and we learn what is expected of individuals belonging to these categories (Bigler & Liben, 2006; Holtgraves & Kashima, 2007; Thompson & Fine, 1999). Communication is the main engine that drives the emergence and maintenance of consensually shared category knowledge. In this process, language plays a crucial role. When communicating about other people and their behavior, our language echoes the existing stereotypic expectancies we have with categorized individuals (often without our conscious awareness). Likewise, when communicating with individuals we have categorized in a given social category, our language subtly reveals the stereotypic expectancies we have about our conversation partner. Research on linguistic bias has revealed a number of subtle systematic variations in language use that reflect stereotypic expectancies in our communications, and as a result, strengthen them in both sender and recipients. This chapter provides an overview of the most important research findings on this topic and places these in an intergroup context.

Linguistic Bias in Communications About Categorized Individuals

When communicating about other people and their behavior, our language echoes the (shared) stereotypic representations of any activated social categories we (implicitly) associate with these people. Stereotypic beliefs surface in (often subtle and largely implicit) linguistic biases, by
means of which existing stereotypic beliefs are shared and confirmed. A linguistic bias can be defined as a systematic asymmetry in word choice that reflects the social category cognitions that are applied to the described group or individual(s).

Three types of biases are distinguished in the literature that reveal, and thereby maintain, social category knowledge and stereotypes: (a) biases in labeling, (b) bias in what we communicate about, and (c) biases in how we formulate information about categorized individuals.

Bias in Labeling

Labeling refers to using specific word(s) to denote a social category. The labels we use reflect which aggregates of people are singled out as meaningful categories. By using conventional category labels, people continuously confirm and maintain which categories are considered as meaningful for categorization within a (sub)culture. Cultures commonly have category labels for age groups (e.g., elderly, adolescents), for ethnicity (e.g., Asian, Caucasian), nationality (e.g., Canadian, Dutch), gender (e.g., men, women), and professions (e.g., plumber, professor).

There are, however, many arbitrary characteristics based on which we could—but conventionally do not—categorize aggregates of people. We do not, for instance, have a label for people who regularly eat carrots, or people who wear colored (vs. black) shoe laces. However, such unconventional categorizations may suddenly become relevant in particular circumstances (e.g., carrots turn out to be poisonous, or shoe laces become a fashion statement). In such circumstances, the group may become a topic of discussion, and people will likely develop a category label to refer to them. In communications, they will likely initially be described (e.g., “people who regularly eat a carrot”), but soon a more definite noun label will evolve (e.g., carrot-eaters). After an initial categorization, subsequent use of pronouns (e.g., we, us vs. they, them) can distinguish individuals as members of such an identifiable out-group.

The use of category labels has important consequences for impression formation. Once a group is linguistically labeled, it is explicitly defined and distinguished from other groups, and thereby gains in apparent reality. Several studies have shown that even trivial category labels induce perceivers to accentuate perceived similarity among members within the labeled category (i.e., they are all alike) and to exaggerate the differences between categories (e.g., Corneille & Judd, 1999; McGarty & Turner, 1992). Foroni and Rothbart (2011, 2013), for instance, presented participants with silhouette drawings of female body types, ordered on a continuum from very thin to very heavy, and had participants estimate their absolute weight and similarity in “personality,” “life style,” and “body type.” The silhouettes were presented in various conditions, in visual groupings, and either with or without a label. Results showed that the presence of a label (compared to no label), even when participants self-generated these, reduced perceived differences between members of the same (labeled) category, while the perceived differences between members of different categories became larger. This shows that once a label is imposed on an aggregate of individuals it obscures our perception of diversity between individual category members.
Moreover, the formation of a stereotype becomes more likely when a category label is used. That is, the more a collection of individuals is perceived as a meaningful, coherent group, the more likely perceivers will seek stereotypic characteristics that are considered to be essential to its members (Abelson, Dasgupta, Park, & Banaji, 1998; Crawford, Sherman, & Hamilton, 2002; Hamilton & Sherman, 1996; Yzerbyt, Corneille, & Estrada, 2001). Once such a stereotype is formed, the use of the category label will be enough for recipients to bring this existing stereotype to mind (Dijksterhuis & Van Knippenberg, 1996).

Interestingly, however, different types of linguistic labels are associated with a different strength in recipient inferences. When individuals are labeled with a noun (e.g., “a carrot-eater,” “a chocolate-eater”), recipients infer that the relevant characteristic is more essential, stable, and resilient to the person, compared to when the individual’s characteristic is described by verbs (e.g., “eats carrots whenever she can,” “eats chocolate a lot”; Gelman & Heyman, 1999; Walton & Banaji, 2004).

A similar difference in recipient inferences has been shown for even more similar grammatical forms, namely nouns versus adjectives. Nouns and corresponding adjectives can be very similar in form (e.g., being a German vs. being German; a homosexual vs. homosexual) as well as in function (e.g., ascribing particular characteristics, such as nationality; Graf, Bilewicz, Finell, & Geschke, 2013). Nevertheless, Carnaghi, Maass, Gresta, Bianchi, Cadinu, and Arcuri, (2008) showed that nouns have a more powerful impact on impression formation than adjectives. Nouns, such as “an athlete,” more strongly induce stereotype-congruent inferences about the target, and simultaneously inhibit counter-stereotypical inferences compared to adjectives, such as “athletic.” Furthermore, nouns induce stronger dispositional inferences compared to adjectives. That is, the relevant characteristic (i.e., being athletic) is seen as part of a profound and unchangeable disposition, as more informative about the person, to have a higher enduringness, and higher likelihood of future repetition (Carnaghi et al., 2008). A noun category label, compared to an adjective, thus conveys more stereotype confirming inferences. Moreover, when in- and out-group members are labeled with nouns (compared to adjectives), this induces a stronger intergroup bias (i.e., ingroup-favoritism; Graf et al., 2013).

Another important difference between label types is whether they are formulated with a plural noun (e.g., Germans are . . . ) or a singular noun (e.g., This German is . . . ). Plural nouns typically have a generic reference, in that they refer to the category as a whole. By contrast, singular nouns single out an individual group member. Plural, compared to specific singular labels, induce stronger recipient inferences in both children and adults (Rhodes, Leslie, & Tworek, 2012). That is, when a new category is referred to using a plural noun (or using indefinite singular generics that have a generic reference, e.g., The German is . . . ), the associated properties are expected to be innate and inevitable and to generalize to other category members. A singular label still induces stronger inferences compared to no label, but the use of a plural label leads to the strongest category expectancies. Other research in developmental psychology with pre-school children confirms the importance of plural noun statements (e.g., dogs are friendly) in the formation of (social) category stereotypes (Bigler & Liben, 2006; Cimpian & Markman, 2009).
Importantly, the use of these different types of labels depends on the information that is communicated. The use of strong definite noun labels (e.g., a Muslim) is reserved for people showing characteristics and behavior that fits with existing stereotype expectancies for the applied category (e.g., regularly visits mosque; fasts during Ramadan). In contrast, to refer to people showing characteristics or behaviors that are inconsistent with existing social category expectancies, speakers are less likely to use a strong noun category label. Individuals showing stereotype inconsistent characteristics and behavior are more likely referred to using modified noun labels (i.e., subtypes; non-practicing Muslim), or with adjectives (e.g., Muslim person), or instead categorized in an alternative category (e.g., an Arab; Richards & Hewstone, 2001). Likewise, generic noun labels are particularly used when speakers believe the category is a distinct group rather than a non-distinct kind (Rhodes et al., 2012).

In some cases, speakers can use different category labels to refer to the same social group. Different labels, even when referring to the same group, may become associated with a different stereotype content (Maass, Suitner, & Merkel, 2014). Derogatory group labels or stigma, for instance, are associated with a different, more negative representation than neutral labels (Carnaghi & Maass, 2007; Dovidio, Major, & Crocker, 2000). When negative and derogatory (metaphorical) nouns are used in this way to describe certain social groups, this can have disastrous, negative consequences for the group described. Musolff (2014), for instance, describes how the use of a “parasite” label to refer to immigrants has legitimized negative and discriminatory language towards this specific (minority) group. The type of label used may call for and justify particular responses or even policies dealing with individuals from the category (Moncrieffe & Eyben, 2013).

In sum, any (sub)culture develops its own conventional labels for (social) categories. In some cases, speakers can use different category labels to refer to the same social group. The choice of one of these labels over the other implies different characteristics of the described group or individuals. Research on linguistic biases shows that, even within the choice of one specific label, different linguistic choices lead to different inferences about the described group. That is, using a noun label over an adjective label leads to stronger inferences in recipients. The use of strong definite noun labels, however, is reserved for people showing stereotype consistent characteristics and behaviors for the applied category. To refer to stereotype inconsistent individuals, in contrast, people tend to use labels that induce weaker stereotype inferences (e.g., adjective labels, or nouns modified by an adjective indicating subtypes). Thus, the choice of one label over the other not only follows from, but also induces recipient inferences that are in line with the social category stereotype of those using the label.

Bias in What We Communicate About Categorized Individuals

A second type of bias is related to stereotype content and deals with the question of what information is shared about people once they have been categorized. Numerous studies on serial reproduction and dyadic conversation have revealed a stereotype-consistency bias. This shows that stereotype-consistent information, compared to stereotype-inconsistent information, is more prominent in communications about categorized individuals (e.g., Kashima, 2000; Klein,
Once target individuals are labeled as a member of a specific category (e.g., by using a label like “an alcoholic”), conversation dyads increase their focus on label-congruent information (Ruscher, 1998; Ruscher, Hammer, & Hammer, 1996). Particularly when information is already part of the common ground (i.e., shared cognition) between communication partners, communication of stereotype-congruent knowledge becomes more likely (Fast, Heath, & Wu, 2009; Lyons & Kashima, 2001, 2003).

Such a general tendency for people to systematically communicate more about certain aspects and traits of categories relative to others has the potential to determine stereotype content (Schaller et al., 2002). Members of (sub)cultures are exposed relatively more to congruent information of shared stereotypes, which leads to a continuous confirmation of existing stereotypic associations. The more often a link between a social category and a certain attribute is perceived, the stronger the attribute in question becomes associated with the social category. This, in turn, increases accessibility and likelihood of being activated and shared again in subsequent communications. By repeatedly retelling a certain stereotypic representation, the representation can thus become a widely shared, cultural belief among members of a subculture (Bratanova & Kashima, 2014).

Next to increased accessibility of existing stereotype content, another important explanation put forward for the tendency to predominantly communicate shared stereotype-consistent information, is that sharing this information is relationally beneficial to communication partners. People prefer to communicate information they believe resonates with their audiences, because it allows them to develop common ground and it facilitates positive perceptions among group members, such as feelings of similarity, liking, and agreeableness (Bratanova & Kashima, 2014; Clark & Kashima, 2007; Higgins, 1992). In addition, conveying shared stereotype-consistent information to a conversation partner likely allows for a relatively smooth interaction. That is, sharing stereotype-consistent information usually requires fewer resources (in communication time, number, and length of utterances) than stereotype-inconsistent information, because conversation partners can easily reconcile the shared information with existing stereotypic beliefs. Furthermore, sharing stereotype-consistent information less likely leads to misunderstanding or disagreement (Klein et al., 2008).

In sum, what is shared about a social category in general, and about individuals who are categorized in a given category, is an important determinant of stereotype content. The tendency to predominantly share information that is part of existing stereotypes is an important factor in the maintenance of stereotypes.

**Bias in How We Formulate Stereotype-Related Information**

A third type of bias is related to linguistic variations in communications about categorized people. When speakers describe the behaviors of categorized individuals, their formulations subtly refer, in various ways, to whether or not the information was stereotypically expected. Research on linguistic biases shows that behavioral information that is in line with social-category knowledge
(i.e., stereotype-consistent) is formulated differently compared to incongruent information (i.e., stereotype-inconsistent). Several subtle variations in verbal formulation can be used to communicate whether a behavior was expected or unexpected for a categorized target (Beukeboom, 2014).

Any language contains an abundance of potential explicit words to indicate that a behavior is typical, or rather that it is an unexpected one-time-event. Such words include adjectives (e.g., expected, invariable, unexpected, surprising), adverbs (e.g., always, constantly, invariably, indeed, unexpectedly, once), determiners (e.g., every, few) and set phrases (e.g., at all times, as always). These words and phrases explicitly refer to existing expectations about the target’s behavior that were either confirmed or challenged.

Language also has more implicit ways to convey existing expectancies. One of the most studied linguistic means associated with stereotype expectedness is language abstraction, as defined by the Linguistic Category Model (LCM; Semin, 2011; Semin & Fiedler, 1988, 1992; Wigboldus & Douglas, 2007). The LCM distinguishes five different word categories that vary on a concrete–abstract dimension. Most concrete are descriptive action verbs (dav). These describe single, observable actions (e.g., “Jack winks at Sue”) and preserve perceptual features of the event. The second category (interpretive action verbs and state action verbs) describes specific observable events, yet are more abstract in that they refer to a general class of behaviors and do not preserve the perceptual features of an action (e.g., “Jack flirts with Sue”). The next category (state verbs) typically describes an unobservable emotional state and not a specific event (e.g., “Jack likes Sue”). Most abstract are adjectives (e.g., “Jack is flirtatious”) that describe only the subject, show no reference to context or specific acts and thus generalize across specific events and objects. The level of abstraction in descriptions of social events shapes the types of inferences that recipients draw. With increasing abstraction, the more recipients infer that the behavior was due to stable and resilient dispositions, rather than contextual and situational forces, and the more likely it is to recur in the future (Semin & Fiedler, 1988, 1992).

The LCM formed the basis for a major contribution to a linguistic mechanism underlying the communication of stereotypes; the Linguistic Intergroup Bias (LIB; Maass, Salvi, Arcuri, & Semin, 1989). The LIB refers to the hypothesis that desirable behaviors of in-group members (i.e., members of the same group the speaker belongs to) and undesirable behaviors of out-group members (i.e., members of a rival group compared to the speaker’s group) are described at a relatively high level of language abstraction (e.g., adj; the in-group member is helpful; the out-group member is aggressive). In contrast, to describe an out-group member showing desirable behavior and an in-group member showing undesirable behavior, relatively low levels of language abstraction are used (e.g., dav; the in-group member hits somebody; the out-group member opens the door for someone; Maass et al., 1989). Because the different LCM categories elicit different cognitive inferences, the implicit meaning that is communicated varies as a function of level of abstraction. By describing desirable behavior of in-group members and undesirable behavior of out-group members abstractly, these behaviors are portrayed as stable and highly diagnostic traits. Undesirable behavior of in-group members and desirable behavior of out-group members, in contrast, are portrayed as situationally determined and exceptions to the rule.
Linguistic Bias

Research demonstrated that the LIB mechanism also operates outside an intergroup context and may result from general expectancies (Maass, Ceccarelli, & Rudin, 1996; Maass, Milesi, Zabbini, & Stahlberg, 1995). Given that expected behavior is considered to be more stable, diagnostic, and typical than unexpected behavior, expected behavior is more appropriately described with abstract terms. Wigboldus, Semin, and Spears (2000) demonstrated that stereotypic expectancies give rise to differences in language abstraction and termed this phenomenon the Linguistic Expectancy Bias (LEB). For example, to describe behavior that is inconsistent with the male stereotype (e.g., crying), people use relatively more concrete language (e.g., he has tears in his eyes). In contrast, when describing a woman demonstrating the same—but female stereotype-consistent—behavior, people tend to use more abstract language (e.g., she is emotional).

Given that abstract descriptions provide more information about the actor’s stable dispositional qualities and less information about the specific situation (Semin & Fiedler, 1988), abstract descriptions of stereotype-consistent behavior endorse existing stereotypic beliefs. That is, abstract words used for stereotype-consistent behaviors imply attribution to stable traits that are likely to be repeated (internal attribution), whereas concrete words used for stereotype-inconsistent behaviors imply attribution to transient situational causes, suggesting exceptions to the rule (external attribution; Wigboldus et al., 2000; Wigboldus & Douglas, 2007). As recipients infer this information from these subtle variations in behavior descriptions, this endorses the stereotypic expectancies the speaker conveys in his/her language.

Besides language abstraction, research revealed a number of other linguistic means that function to maintain stereotypes. One way by which speakers convey that a person’s behavior is stereotype inconsistent is shown in the Stereotypic Explanatory Bias (SEB; Hamilton, & Sherman, 1996; Hastie, 1984; Sekaquaptewa, Espinoza, Thompson, Vargas, & von Hippel, 2003). The SEB proposes that speakers tend to produce more explanatory comments for stereotype-inconsistent (vs. consistent) behaviors. Such explanations of stereotype-inconsistent behavior are argued to reflect a speaker’s attempt to make sense of a surprising inconsistency (e.g., the man was crying, because he had a rough day). For stereotype-consistent behavior (e.g., the woman was crying), in contrast, such an explanation is not called for. The explanations thus reflect a speaker’s attempt to resolve why the inconsistency occurred in order to maintain coherence in one’s existing stereotypic impression. Both concrete situational descriptions and explanations are considered to be linguistic reflections of such cognitive attempts to resolve inconsistencies.

When communicating about stereotype-inconsistent information, speakers can also introduce stereotype-consistent terms, for instance by using negations or irony. Research on the Negation Bias (NB; Beukeboom, Finkenauer, & Wigboldus, 2010) revealed that the use of syntactic negations (e.g., not stupid, rather than smart) is more pronounced in descriptions of stereotype-inconsistent compared to stereotype-consistent behaviors. For example, if one holds a stereotypic expectancy that girls are emotional, but a particular girl violates this expectancy by being very tough, a speaker is likely to reveal his prior expectancy by using a negation like The girl was not emotional. In contrast, for stereotype-consistent behavior (e.g., the boy was tough; the girl was emotional), the use of negations is less likely. Negations thus allow one to introduce stereotype-consistent concepts in communications about stereotype-inconsistent information,
and thereby to re-affirm existing associations with a category. Because negations tend to activate rather than suppress the negated concept, negations likely reinforce the association between the target and the negated concept (e.g., girls are emotional) in a recipient (Beukeboom et al., 2010).

A similar mechanism occurs in the Irony Bias (IB; Burgers & Beukeboom, 2016). Research on the IB shows that speakers find ironic remarks particularly appropriate to comment on stereotype-inconsistent (vs. stereotype-consistent) behaviors. An ironic remark about stereotype inconsistent behavior (e.g., “He really is a tough guy” about a man who is crying) allows speakers to introduce the expectancy (i.e., guys are tough) and also indicate that the expectancy is disconfirmed in the given situation. Like negations, ironic remarks introduce opposite terms to describe a behavioral situation and allow speakers to refer to implicit stereotypes. Both negations and ironic comments about stereotype-violating behavior can thus activate and communicate the implicit stereotypic expectancy in message recipients, even though the actual situation is stereotype disconfirming.

Moreover, recipients infer that a behavior is unexpected for a described person when it is described with a negation (compared to an affirmation; Beukeboom et al., 2010) or using ironic (compared to literal) comments (Burgers & Beukeboom, 2016). Just like the LEB and SEB, these biases thus serve to maintain knowledge about what is and what is not expected for members of a given social category. Even when people are confronted with stereotype-inconsistent information, a biased description may induce recipients to draw stereotype-confirming inferences.

**Linguistic Bias in Communications With Categorized Individuals**

Another area of research focuses on how stereotypic expectancies influence our communication with categorized individuals. When interacting with others, people draw all kinds of inferences about their conversation partner’s thoughts, goals, intentions, and attitudes. Such inferences play a crucial role in the coordination of the conversation and are needed to obtain mutual understanding (Krauss & Fussell, 1996). Importantly, the inferences we draw will be partly based on activated stereotypes that are applied to our conversation partner. We will have different expectancies about a conversation partner of age 5 versus age 85, or a person categorized as a plumber versus a professor, and we change our behaviour accordingly. Research in the area of Communication Accommodation Theory (Dragojevic, Gasiorek, & Giles, 2015) and interracial (Trawalter, Richeson, & Shelton, 2009) and intergenerational (Hummert, Garstka, Ryan, & Bonnesen, 2004) communication, for instance, shows how activated stereotypes about our conversation partners influence our communicative behaviors.

The linguistic biases described in the previous sections have hardly been explicitly related to this area of research. However, when we provide someone with feedback about their behavior, our language likely reveals our existing expectancies about the described behavior in a similar manner as described in the previous section. The use of negations, for instance, can reveal a speaker’s prior negative expectancy (e.g., “Not bad,” “You did not make many mistakes”; cf. Beukeboom et al., 2010). Note that situations in which we discuss the behaviors of our
conversation partners are common, for instance, in education (teacher–pupil conversations), and in appraisal interviews in a professional context. Compared to communicating about an absent target, it is probable that other factors come into play when communicating with a target who is also the recipient of a description. Speakers likely take the target’s feelings into account, and utterances will more likely focus on the individual rather than a category as a whole. However, as indicated above, speakers are often unaware that they reveal stereotypic expectancies in their formulations, and this is likely also the case when they talk to categorized individuals.

Linguistic biases in communication towards members of minority groups (e.g., based on race, gender, sexual orientation) have been described as micro-aggressions. Micro-aggressions are defined as subtle insults directed toward a person who threatens and demeans the target (Sue, 2010). While the people perpetrating them are usually unaware they are causing harm and often intend no offense, targets may be sensitive to the subtleties in language that reveal they are being categorized and associated with negative stereotypic traits.

Being categorized—albeit by means of very subtle linguistic cues—may have several serious effects on targets. It may induce them to confirm expectancies conveyed by the speaker as self-fulfilling prophecy (Hummert et al., 2004; Word, Zanna, & Cooper, 1974); it can induce impaired performance as a result of stereotype threat (Shapiro & Neuberg, 2007; Steele, 1997) or improved performance as a result of stereotype lift (Walton & Cohen, 2003) and can result in lower self-esteem (Bourguignon, Yzerbyt, Teixeira, & Herman, 2015) and even deteriorate mental and physical health (Dovidio et al., 2000; Pascoe & Richman, 2009). Thus, even though micro-aggressions typically are not classified as explicit cases of racism and/or prejudice, they can still have a profound negative impact on the people targeted.

Discussion and Future Research Directions

In sum, when we refer to categorized people and their behaviors, or when we talk to people we have categorized, our stereotypic expectancies are often reflected in our language use. The types of category labels we use reflect existing social category cognitions, and once a category is activated, we tend to predominantly communicate stereotype-consistent (rather than inconsistent) information. If inconsistent information is introduced in discourse, it tends to be subtyped, and/or explained (SEB), and/or framed as a transient one-time situation rather than caused by stable dispositional factors (LIB, LEB), and/or related to stereotype consistent term by means of negations or irony (NB, IB). In contrast, consistent information is framed such that it allows one to reconfirm existing stereotypes, by using strong category labels (nouns) and more abstract language (LIB, LEB) that implies stability and likelihood of recurrence.

Much of the discussed biases in language use are considered to operate largely outside of people’s awareness. People make millions of linguistic choices every day, and the reflection of stereotypic beliefs in language use typically occurs unintentionally (Franco & Maass, 1996; Maass, 1999). The words that people choose when describing or addressing individuals belonging to different social categories unintentionally reflects existing social category knowledge and cognitive attempts to make sense of observed inconsistencies with existing stereotypic knowledge (Beukeboom, 2014).
These effects appear to be difficult to inhibit (Franco & Maass, 1996), and some biases have even been used as implicit measures of prejudice (i.e., LIB/LEB, Von Hippel, Sekaquaptewa, & Vargas, 1997; and SEB, Sekaquaptewa et al., 2003).

It is important to note a number of factors that can moderate the reflection of stereotypes in biased language use. First, our utterances obviously do not merely result from privately held intra-personal cognitions, like stereotypes. Depending on the communicative situation, our utterances are often tailored to suit interpersonal communication goals (Higgins, 1992). The characteristics of recipients, and the sender and recipient’s relation to each other and/or the target, will determine whether linguistic biases occur (Freytag, 2008). As noted, the biases we discussed here are mainly expected when sender and recipient share common ground about currently relevant and activated social-category cognition. Several authors have argued that sharing social category knowledge and stereotypes through communication is a form of communicational grounding (Beukeboom, 2014; Fiedler, Bluemke, Friese, & Hofmann, 2003; Klein, Clark, & Lyons, 2010). The sharing of category knowledge should thus be heightened in subcultures that already share stereotypic knowledge (Clark & Kashima, 2007; Fast et al., 2009).

Second, the communicative situation may evoke communication goals that have a strong effect on our language use. In situations in which speaker and recipient do not share common ground, the speaker, for instance, may become motivated to explain something or convince a recipient (Fiedler et al., 2003). Such contextually induced communication goals may overrule effects of activated stereotypic expectancies (Douglas & Sutton, 2003). When motivated to favorably portray a person or group, for instance, speakers can adopt abstract predicates to describe positive behaviors and concrete predicates to describe negative behaviors (cf. LIB/LEB). Douglas and Sutton (2003) showed that communication goals have a strong effect on the use of language abstraction, and can, consequently, completely reverse linguistic biases.

Another important motivational factor that affects language use arises from a speaker’s social identity. The key insight of Social Identity Theory (Tajfel & Turner, 1979) and Self-Categorization Theory (Hogg & Reid, 2006; Turner et al., 1987) is that we not only categorize others, but also categorize ourselves as members of social groups. These self-categorizations are an inherent aspect of people’s identity, and people strive to maintain and enhance a positive social identity. This desire can be reflected in the types of comments speakers make about behaviors by in- and out-group members, as is shown in the LIB in the use of predicates of different abstraction.

Research on the underlying mechanisms of the LIB demonstrated that referring to positive and negative behaviors of in- and out-group members is driven by a motivation to protect one’s social identity (Maass et al., 1995, 1996; Wigboldus & Douglas, 2007). It was demonstrated that the LIB was more pronounced in intergroup settings wherein the ingroup was threatened (i.e., hostility between Northern and Southern Italians). This motivational effect that results from a speaker’s group memberships, was shown to be independent of the expectancy mechanism by which implicit cognitive associations and expectancies are reflected in language abstraction; i.e., behaviors that are expected (vs. unexpected) for Northern and Southern Italians are described at a
higher level of abstraction (Maass et al., 1995; Maas, Ceccarelli, & Rudin, 1996) As described above, further research confirmed this latter mechanism in research on the LEB (Karpinski & Von Hippel, 1996; Wigboldus et al., 2000).

Although most research on these mechanisms has been conducted on language abstraction, it seems reasonable to assume that other linguistic biases could be driven by similar motivations. Burgers, Beukeboom, Kelder, & Peeters (2015), for instance, showed how soccer fans can employ ironic remarks to enhance group identity. Irony about competent and incompetent behaviors of in-group and out-group team players can be used both as a linguistic tool for aggression towards out-group members and to subtly communicate expectancies about desired in-group and out-group behavior.

While much progress has recently been made in uncovering the ways in which language shapes and maintains stereotypes, many opportunities for future research remain. The language types described (e.g., nouns, adjectives, verbs; language abstraction, negations, irony) constitute but a small part of the choices that language users can make. Undoubtedly there are more biases that convey stereotypic expectancies in language that have hitherto not been studied. Future research in the area of linguistic bias could turn towards research in theoretical and corpus linguistics to identify potential other linguistic variables that may operate as a source of linguistic bias.

Such avenues could both broaden and narrow the focus for linguistic-bias research. In terms of broadening, most linguistic-bias research has currently focused on linguistic properties associated with parts of speech (by contrasting verbs, adjectives, nouns) and pragmatics (through the LCM and the focus on irony). Yet, linguistics has identified multiple other sub-fields such as semantics, morphology, and phonetics that may all play a role and interact in the formation and maintenance of social stereotypes. As such, the linguistics literature may work like a treasure trove to further specify which kinds of linguistic variables function in which ways in forming and/or maintaining stereotypes.

In terms of narrowing, the linguistics literature can also help to further specify how linguistic variables operate in stereotype communication. For instance, the NB proposes that negations play an important role in stereotype communication (Beukeboom et al., 2010). However, empirical evidence for the NB is mostly based on research on syntactic negations (e.g., no, not). The linguistics literature proposes that syntactic negations are but one of several forms of negations (e.g., Verhagen, 2005). Another type of negation is a morphological negation, in which a word is negated through a morphological prefix or suffix that is attached to a word (e.g., the prefix “un” in a word like unhappy, which can be seen as a negation of happy). These morphological negations are proposed to work differently from syntactic negations, in that morphological negations do not first activate the negated concept. That is, a sentence like “The clown is not happy” is 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, Schul, & Burnstein, 2004). By contrast, a morphological negation in a sentence like “The clown is unhappy” is processed in one step, with addressees immediately coming to the intended meaning of unhappy without first
activating the negated concept (happy). The implications can thus be that the NB primarily works for syntactic (but not for morphological) negations, which is a question to address in future research.

Another interesting avenue to explore lies in the question of whether the described biases are linguistic universals. With the exception of the Linguistic Intergroup Bias, which has been demonstrated in various languages (Maass, 1999) including Chinese (Lu & Zhang, 2014), research on other linguistic biases has been conducted in one or a few languages. For example, current research on the NB and the IB has primarily focused on Dutch (Beukeboom et al., 2010; Burgers & Beukeboom, 2016). Future research can explore whether the predictions of the NB and the IB also hold true in languages that are markedly different from a Germanic language like Dutch, such as Uralic languages like Hungarian and Finnish, Sino-Tibetan languages like Mandarin Chinese, or Japonic languages like Japanese. Such cross-linguistic research can reveal if speakers of different language types form and maintain stereotypes in similar or different ways.

Finally, most research on linguistic bias has been conducted in experimental settings, mostly by manipulating artificial sentences in isolation. As a next step, it would be interesting to study how stereotypes are communicated in natural language, in which various biases may occur in combination. This would also shed light on the question of whether different biases co-occur or perhaps compensate their functionality (e.g., label types with levels of abstraction and/or affirmations vs. negations). Studying biased language use in spontaneous conversations, newspaper articles, or social media interactions, for instance, will provide a richer understanding of how stereotypes are maintained in real life.

To conclude, biased language use is not necessarily harmful, nor evidence of discrimination and prejudice. Biased language may simply result from a tendency to find common ground with one’s communication partner; and sharing stereotypes is, given their functional role, highly valuable. Nevertheless, it is important to be aware of stereotype-maintaining biases as they play a crucial role in consensualizing both benevolent and harmful stereotypes about social categories. The study of linguistic bias can reveal in which ways our verbal communications can be a help or a hindrance in forming, perpetuating, or challenging stereotypic views.

Further Reading


**References**


Linguistic Bias


### Related Articles

- Prejudiced Communication
- Interethnic Communication
- Gender Bias and Sexism in Language
- Detection of Deception
- The Politics of Translation and Interpretation in International Communication