A DEVELOPMENTAL PERSPECTIVE
ON LINGUISTIC INTERGROUP BIASES

A THESIS
Submitted In Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

By
Mirja Otten
University of Hamburg
October 2021
Chairman of the Committee: Prof. Dr. Alexander Redlich
   Professor of Educational Psychology
   University of Hamburg
Advisor and Primary Reviewer: Degner, Juliane, Prof. Dr.
   Professor of Social Psychology
   University of Hamburg
Secondary Reviewer: Wagner, Jenny, Prof. Dr.
   Professor of Educational Psychology and Personality Development
   University of Hamburg
Committee Member: Oettingen, Gabriele, Prof. Dr.
   Professor of Educational Psychology and Motivation
   University of Hamburg and New York University
Committee Member: Lincoln, Tania, Prof. Dr.
   Professor of Clinical Psychology and Psychotherapy
   University of Hamburg

Date of Disputation: April 20th, 2022
ABSTRACT

The linguistic intergroup bias (LIB) describes systematic differences in language use that are related to underlying intergroup attitudes: Individuals tend to describe desirable behaviors of ingroup members in abstract linguistic terms and undesirable behaviors of ingroup members in more concrete terms. For reports on behavior of outgroup members, this pattern reverses.

This difference signals different perceptions and representations of in- and outgroups: On the one hand, positive ingroup behavior and negative outgroup behavior are perceived as more diagnostic about the actors and their social group. On the other hand, negative ingroup behavior and positive outgroup behavior are represented as rather rare events.

Such subtle differences may not only indicate existing group representations, but also facilitate the social transmission of stereotypes. It has been assumed that biased language is an important factor in the socialization of stereotypical thinking. However, empirical support for this assumption is scarce and to this point we lack a clear developmental perspective on the association of language abstraction and attitude formation. Following the assumption that there is a causal effect of language abstraction on impression formation respectively formation of group evaluation, this effect should be visible already early in childhood, since empirical evidence has placed emergence of intergroup attitudes in this developmental phase.

In a series of studies, we presented children with storybooks in which agents of two novel social groups demonstrate desired and undesired behaviors. In each study, we manipulated with which level of abstraction group behaviors were described. We found however no systematic effect of language abstraction on group evaluation, neither for novel groups that were presented
with unambiguous behavior, i.e., behaving either only positively or only negatively, nor for groups presented with ambiguous behavior, i.e., groups behaving equally positive and negative. To investigate whether a potential effect of language abstraction is subject to a certain developmental level, the sample of one of our experiments additionally includes adults, yet we found no systematic effect of language abstraction on group evaluation for adults either.

We discuss the potential effects of language abstraction on impression formation, acknowledging the experiments’ limitations and drawing an outline for further research.

Keywords: Linguistic intergroup bias, language abstraction, attitude formation, intergroup attitudes
ACKNOWLEDGMENTS

First and foremost, I would like to thank my advisor Prof. Dr. Juliane Degner, who inspired this line of research. Throughout the years working in Hamburg, Juliane provided me with a truly supporting working environment. I am very thankful for her encouragement, theoretical as well as methodical advice and a laugh when needed.

Likewise, I am very thankful to Prof. Dr. Jenny Wagner for agreeing to be my second supervisor.

Thanks are also due to Prof. Dr. Tania Lincoln and Prof. Dr. Gabriele Oettingen who agreed to take part in the dissertation committee.

As all empirical research projects, this thesis was in need of cooperation partners for data collection. Without the consent of numerous parents, the cheerful participation of so many children, the helpfulness of principals, teachers and educators, this thesis would not have been possible. I owe special thanks to all of them. Collecting data in kindergarten and schools was a rewarding and amusing but sometimes lengthy and challenging process. In the course of this project a number of students, students’ assistants and interns supported this project by assisting with data collection. For their help and sometimes inconveniences, I want to say thanks to Jana Mangels, Marc Rommel, Habiba Schiller, Lena Brachlow, Katja Coser, Maibrit Arbien, Karin Reisinger, Anna Maxhofer and Catherine Maschler and especially to Charlotte Aline Fischer.

Finally, heartfelt thanks go to my understanding and loving family, that gifted me with a life full of opportunities; to all my friends, especially Marilena, for listening to me with patience when I was about to lose patience for my project and for encouragement and motivation.
### TABLE OF CONTENTS

**ABSTRACT** .......................................................... iv

**ACKNOWLEDGMENTS** ............................................... vi

**LIST OF TABLES** ..................................................... xi

**LIST OF FIGURES** ................................................... xii

**Introduction** ........................................................................ 1

**A Developmental Perspective on Stereotype Formation** .................... 5

- Social-Cognitive-Developmental-Theory ........................................ 5
- Social Identity Development Theory ............................................... 6
- Developmental Intergroup Theory .................................................. 7

**Conclusions** ........................................................................ 10

**How Words Represent Impressions of Inner Worlds: The Linguistic Category**

**Model** ........................................................................... 10

**How Words Represents Biases in Inner Worlds: The Linguistic Intergroup Bias**... 19

**Cognition or Motivation – What drives the Linguistic Intergroup Bias?** ............. 22

**Linguistic Biases and Stereotypic Beliefs: A Hen and Egg Problem** ................... 28

**Linguistic Intergroup Bias from a Developmental Perspective** ....................... 32

**OUTLINE of the present research project** ............................................ 38

- Linguistic Manipulation of Group Perception for positive or negative novel groups
  
  **(Experiment 1)** ...................................................................... 41

- **Method** ........................................................................... 41

- **Materials** .......................................................................... 41
Experiment 2: Effects of Linguistic Differences in Behavioral Descriptions on the Acquisition of Intergroup Evaluations by Preschool Children .......................... 66
DEVELOPMENTAL PERSPECTIVE ON LIB

Group evaluation...........................................................................................................80
Group preference...........................................................................................................85
Reward Allocation..........................................................................................................85
New member evaluation...............................................................................................87
New member preference...............................................................................................90
Ambiguous behavior task.............................................................................................90
Discussion......................................................................................................................91

Linguistic Manipulation of Group Perception for positive or negative novel groups

(Experiment 3a).............................................................................................................96

Method...........................................................................................................................98
Materials..........................................................................................................................98
Participants......................................................................................................................99
Dependent Variables.....................................................................................................100
Procedure.......................................................................................................................102
Design.............................................................................................................................103
Hypothesis......................................................................................................................104
Results.............................................................................................................................105

Group evaluation...........................................................................................................105
Group preference...........................................................................................................109
Reward Allocation..........................................................................................................110
New Member Evaluation...............................................................................................111
New Member Preference...............................................................................................115
Ambiguous Behavior Task............................................................................................116
APPENDIX C: Example of AN AMBIGUOUS BEHAVIOR TASK

LIST OF TABLES

Table 1. Classification Criteria for linguistic category as suggested by Semin & Fiedler (1988) and in an extension by Carnaghi et al. (2008) .......................................................... 16

Table 2. Example of linguistic biases in description of positive and negative behaviors ........ 20

Table 3. Reliability of the positive and negative evaluation scales for group evaluation in Experiment 1 .......................................................... 80

Table 4. Mean positive and negative evaluations of new members in Experiment 1 .......... 86

Table 5. Reliability of the positive and negative evaluation scales for new member evaluation in Experiment 2 ........................................................................................................ 87

Table 6. Distribution of participants per sample and condition in Experiment 3a ........... 99

Table 7. Reliability of the positive and negative evaluation scales for group evaluation in Experiment 3a .......................................................... 107

Table 8. Interaction between group and item valence for each condition and sample combination of Experiment 3a .......................................................... 108

Table 9. Main effect of group membership for reward allocation in all experimental conditions of Experiment 3a .......................................................... 110

Table 10. Reliability of the positive and negative evaluation scales for new member evaluation in Experiment 3a .......................................................... 114
LIST OF FIGURES

Figure 1. Example episode of negative behavior.................................................................12
Figure 2. Example episode of positive behavior.................................................................12
Figure 3. Example cartoon used by Maass et al. (1989), extracted from Maass (1999), p. 82......20
Figure 4. Example episode from the storybook in Experiment 1........................................43
Figure 5. Examples for the scale used to assess dependent variables. ..............................46
Figure 6. Group based trait evaluation in the three experimental conditions of Experiment 1…54
Figure 7. Preference for groups in absolute numbers for the three experimental conditions of
Experiment 1....................................................................................................................56
Figure 8. New member evaluation in the three experimental conditions of Experiment 1.......58
Figure 9. Preference for new members in absolute numbers for the three experimental conditions
of Experiment 1................................................................................................................59
Figure 10. Averaged number of positive interpretations of ambiguous group behaviors........60
Figure 11. Example episode from the storybook in Experiment 2........................................74
Figure 12. Group evaluation in the preschool sample for the three experimental conditions of
Experiment 2....................................................................................................................82
Figure 13. Group evaluation in the elementary school sample for the three experimental
conditions of Experiment 2............................................................................................83
Figure 14. Preference for groups in absolute numbers for the three experimental conditions of
Experiment 2....................................................................................................................84
Figure 15. Distribution of rewards in the three experimental conditions of Experiment 2........85
Figure 16. New member evaluation per sample in the three experimental conditions of Experiment 2. 

Figure 17. Preference for new members in absolute numbers for the three experimental conditions of Experiment 2.

Figure 18. Averaged number of positive interpretations of ambiguous group behaviors in the three experimental conditions of Experiment 2.

Figure 19. Group evaluation in the four experimental conditions of Experiment 3a.

Figure 20. Preference for groups in absolute numbers for the four experimental conditions of Experiment 3a.

Figure 21. Distribution of rewards in the four experimental conditions of Experiment 3a.

Figure 22. New member evaluation in the four experimental conditions of Experiment 3a.

Figure 23. Preference for new members in absolute numbers for the three experimental conditions of Experiment 3a.

Figure 24. Consistent ascriptions of behavior for a) indication of group membership by label only, b) indication of group membership by clothing color only and c) indication of group membership by matching color and label in Experiment 3a.

Figure 25. Group evaluations in the four experimental conditions of Experiment 3b.

Figure 26. Preference for groups in absolute numbers for the four experimental conditions of Experiment 3b.

Figure 27. Distribution of rewards in the four experimental conditions of Experiment 3b.

Figure 28. New member evaluation in the four experimental conditions of Experiment 3b.

Figure 29. Preference for new members in absolute numbers for the four experimental conditions of Experiment 3b.
Figure 30. Consistent ascriptions of behavior for a) indication of group membership by label only, b) indication of group membership by clothing color only and c) indication of group membership by matching color and label in Experiment 3b .........................................................137
Introduction

Have you ever been on your way to work and something happened that really bothered you? Maybe you were hurrying towards a train, and you missed it because someone didn’t hold the door for you - despite seeing you sprinting towards it? Or maybe you were going to work by bicycle, and you almost got hit by a car because the driver was checking his phone and accidentally left his lane?

Let’s assume you experienced the latter, and it bothered you quite enough to talk to someone about it, for example your spouse, a friend, or a colleague, what exactly did you say? Did you tell them that a car nearly hit you? Did you tell them that you almost got hurt? Did you tell them about your encounter with a reckless driver or a road bully?

What kind of description you chose may depend on a variety of factors: your emotionality, your style of narration and a probably less obvious factor: the similarity between you and the person who was driving the car. How so?

If you and the driver belong to the same social category, for example you are both of Caucasian ethnicity, it is more likely that you described the incident with a rather concrete statement, i.e., by saying that a car nearly hit you. However, if you and the driver differ regarding social categories, for example one of you is of Caucasian the other of Arabian ethnicity, it is more likely that you described the incident with a rather abstract statement, i.e., by explaining how reckless the car driver was. Why is that?

While both statements are an equally true description of your experience, they differ regarding an important linguistic feature: the abstractness of your description. That difference is most likely not arbitrary but rather an expression of a typical phenomenon that occurs when we
talk about people with which we share important social categories versus when we talk about people that belong to different social categories. This phenomenon is known as the „Linguistic Intergroup Bias“ (LIB, Maass, Salvi, Arcuri & Semin, 1989).

The LIB is a language pattern in which we use abstract language to communicate about positive behavior of ingroup members and negative behavior of outgroup members, while we use concrete language to describe negative behavior of ingroup members and positive behavior of outgroup members (Maass, 1989). This distinct use of language abstraction is not only a mere linguistic peculiarity but has remarkable impact on impression formation, as for example demonstrated by Semin and Fiedler (1988). Semin and Fiedler categorized different word classes according to their effect on behavioral perception. They demonstrated, for example, that the more abstract a given description of someone’s behavior, the more likely we are to assume the behavior is diagnostic for the person, i.e., the behavior is evaluated as typical for that person and a reflection of specific character traits. Simplified put, there is a link between language abstraction and person perception, and we tend to use our language in a way that forms enduring positive impression of ingroup members and enduring negative impressions of outgroup members. Due to this effect, it has been assumed that biased use of language might affect stereotype perpetuation and transmission not only in adults but children as well.

There are three main theories respectively concepts that are important to analyze when hypothesizing about a causal effect of language abstraction on intergroup attitudes. These are (1) The Developmental Intergroup Theory (DIT, Bigler & Liben, 2006), (2) The Linguistic Category Modell (LCM, Semin & Fiedler, 1988) and the (3) Linguistic Intergroup Theory (LIB, Maass et al., 1989), which were used as a theoretical framework for this thesis.
Integrating postulation of all three makes the conclusion that linguistic abstraction (LCM) in an intergroup context (LIB) contributes to children’s learning (DIT) of rather positive intergroup and rather negative outgroup attitudes at least plausible. To this point, there has however not been empirical evidence for a causal link between language abstraction and impression formation. With this research, we therefore aim to provide empirical evidence on potential effects of language on impression formation, specifically during early childhood. For this purpose, the thesis is structured as follows:

**Presentation of a theoretical framework**

First, we discuss the emergence of intergroup attitudes in more detail and thereby illustrate a possible gateway for an influence of language (abstraction) on children’s impression formation. Second, we introduce the Linguistic Category Model (LCM) to illustrate how language reflects person perception and the Linguistic Group Bias (LIB) to demonstrate how language reflects intergroup attitudes, including considerations on the driving principles of the LIB and possible derivations regarding its functions with respect to the development of intergroup attitudes.

**Deduction of the research question**

We analyze existing research on linguistic biases and demonstrate that there is yet no research clearly providing evidence for a causal effect of language on impression respectively attitude formation, even though there is already research on linguistic biases occurring during childhood and adolescents which will be shortly discussed.

**Experimental investigation and discussion**

We design and conduct a series of experiment to test the assumption that language (abstraction) influences children’s learning and evaluation of novel social groups.
We further discuss the experiments results with regard to the assumption that language abstraction considerably contributes to cross-generational transmission of intergroup attitudes.
A Developmental Perspective on Stereotype Formation

Developmental Psychology has provided several theories on the process of intergroup attitude formation in young children. During the last decades, the Social-Cognitive-Developmental-Theory (Aboud, 2008), the Social Identity Development Theory (Nesdale, 2001) and the Developmental Intergroup Theory (Bigler & Liben, 2006, 2007) turned out to be the most influential developmental theories. While none of these theories explicitly mentions language abstraction as a relevant factor of attitude formation, the authors have partially made presumptions on language-based influences on attitude and stereotype formation.

Hence, an analysis of these theories can help determine which processes of attitude formation might be influenced by language and how language might interplay with other factors relevant for attitude formation in childhood.

Social-Cognitive-Developmental-Theory

The Social-Cognitive-Developmental Theory (SCDT) established by Aboud (1988) is focused solely on the development of ethnic-based prejudice. SCDT directly links prejudice development to general development stages during childhood, thereby substantially building on Piaget’s theory of cognitive development. That is, SCDT assumes that children run through developmental stages associated with specific kind of perspectives (e.g., egocentric or sociocentric) on the self and others, changing levels of (cognitive) abilities and varying predominant psychological processes that result in age-related differences in prejudice expression (Aboud, 2008).

SCDT does hold only scarce assumptions about the origin of intergroup attitudes, but discusses potential socialization influences from children’s social environments. Language is no
The distinct factor in this framework yet regarded a transmitter of information, that provides children with the knowledge bolstering the evaluation of social categories. Thus, Aboud in general acknowledges that possibility that specific linguistic features affect prejudice formation and might explain “why they are differentially adopted by youth” (Aboud, 2005, p. 318).

**Social Identity Development Theory**

Social Identity Development Theory (SIDT) advanced by Nesdale (Nesdale 1999, Nesdale 2001, Nesdale & Flessor, 2001) is essentially based on Social Identity Theory (SIT, Tajfel & Turner, 1978, 1979), i.e., the assumption that people strive for positive self-esteem and partially gain this positive self-esteem through membership in positively evaluated group, and Self-Categorization Theory (SCT, Turner, 1987), i.e., the assumption that people maintain a complex system of self-related beliefs that form their self-concept.

Nesdale suggested that children initially are unresponsive to racial cues, however will start to follow cues given by their social environment in early childhood and thereby develop the ability to differentiate ethnicities, the ability of ethnic self-identification and subsequently ingroup preference will emerge (Nesdale, 2000). Ingroup preference is based on increased liking of the ingroup rather than on increasing dislike of the outgroup. The emergence of prejudice is accordingly not believed to be an automatic consequence of ingroup preference, but a separate process. According to SIDT, whether children actually acquire prejudice depends on three factors, namely ethnic consistency (the belief that ethnic cues do not change with age or circumstances), cognitive skills (i.e. rather low levels of perspective taking, higher moral reasoning) and identification (assimilation of negative outgroup attitudes from their social environment, especially from persons close to them and whom they identify with).
Although SIDT offers no direct references to the role of language, Nesdale discussed how language could contribute to racial division, namely as communicator of racist beliefs. That is, Nesdale assumed an effect of language content rather than of linguistic features of content. Nevertheless, Nesdale has explicitly referred to the phenomenon of the LIB and suggested an examination of its function in the context of prejudice development – hinting at the possibility that, besides the mere content, the ‘how’ of information transmission might as well influence intergroup attitudes.

**Developmental Intergroup Theory**

Developmental Intergroup Theory (DIT; Bigler & Liben, 2006, 2007) is not limited to the explanation of ethnic attitudes but follows a ‘general approach’, meaning that the same variables are assumed to operate similarly – yet with differing strength - in the formation of all kinds of intergroup attitudes, e.g., the same variables and processes are essential for the formation of ethnic as well as gender and other attitudes socially relevant differences.

In the postulation of DIT Bigler and Liben (2006) acknowledge the importance of social as well as cognitive abilities and the reciprocal interaction of individual and environment, thereby partially integrating factors in their theoretical framework that have been outlined by SCDT and SIDT as well. Conclusively, they suggested three steps as crucial regarding intergroup attitude formation. They assumed that (1) specific person-related attributes become psychologically salient and are, thus, established as dimensions for person categorization; (2) children learn to use these attributes for categorization when encountering individuals and (3) children actively construe stereotypes by filling these social categories with meaning. For this latter process of stereotype construction four factors are considered to be of major importance,
these are a) essentialism, b) ingroup bias, c) explicit attributions and d) implicit attributions or group-attribute co-variations (Bigler & Liben, 2006).

The first factor, essentialism, describes the subjective belief that differences between social categories are based on inherent distinct natures, for example, (alleged) biological differences between social groups. Research on essentialism in children demonstrated how the use of abstract language in the form of nouns and generic language contributed to a greater stability of essentialist beliefs. More precisely, conveying essentialist beliefs in abstract terms generated the impression of more stable category memberships and greater stability of personality traits for persons who were subject of these beliefs (Gelman & Heymann, 1999; Gelman, 2004). It is possible that the use of abstract language in form of abstract linguistic categories in a similar way contributes to greater stability of beliefs when attaching meaning to social categories. That is, abstract language in essentialism could cause the construction of more stable stereotypic beliefs and attitudes.

The second factor, ingroup bias, seemingly is an inevitable consequence of social categorization and by that influences the nature of constructed beliefs. As a result of this bias, children are expected to generate more positive beliefs about the ingroup than the outgroup (Dunham, Baron & Carey, 2011; Nesdale, Maass, Griffiths & Durkin, 2003).

The remaining two factors, explicit and implicit attributions, are largely communicated through language and thereby are a carrier of linguistic information. Regarding the examination of linguistic effects on attitude formation, DIT is therefore the most compelling theory as its theoretical framework already contains provisions on the role of language. More importantly, although Bigler and Liben do not refer to the LIB directly their assumptions on potential linguistic effects would allow an integration of effects associated with the LIB.
According to Bigler and Liben language affects attitude formation by explicit attributions that provide a clear association between a social group and a trait or behavior, etc. These associations – and this is where the LIB might take effect – can be presented on rather abstract levels, e.g., “Boys are bullies” or on a rather concrete level “Boys don’t cry”. Depending on the abstraction level used to convey social information, explicit attributions may generate rather stable positive or negative impressions of social groups and by that contribute to the construction of biased intergroup attitudes.

Yet it seems even more likely that subtle linguistic differences assert an influence on stereotype formation via implicit attributions. Bigler and Liben (2006) construed implicit attributions as not language based, but built on nonverbal behavior cues or observations between category membership and specific behaviors or attributes. It is, however, conceivable that linguistic differences in communication are a further source for these implicit attributions – not by the content but by ‘the how’ of communication. When biased language is used for communication about social groups, the underlying differences in linguistics might foster biased inferences about these groups and support the formation of biased intergroup attitudes. Again, it might be the level of abstraction in communication about social groups, more precisely a biased level of abstraction in communication of social groups we belong to and social groups we do not belong to, that significantly contributes to attitude formation. That is, biased intergroup attitudes that are reflected in biased language use of adults might - in terms of the DIT - provide a form of implicit attribution that are suggested to causally influence attitude formation in children.

This reasoning demonstrates that, although DIT does not explicitly account for linguistically biased information, it considers language-based influences on attitude formation that might assert additional influences on attitude formation. Thereby, DIT provides a theoretical
framework that allows the integration of a potential causal effect of language abstraction on attitude formation.

Conclusions

Language is typically not considered an independent factor in theories on developmental intergroup attitudes. DIT is an exception as it explicitly discusses language based effects on attitude formation and moreover offers the most promising possibility to integrate further language-related effects in its theoretical framework, specifically the effect of differing abstraction on communication processes might be easily integrated into the suggested framework – given that there is a causal link between language abstraction and attitude formation in childhood, an assumption which examination is aim of this thesis.

How Words Represent Impressions of Inner Worlds: The Linguistic Category Model

Using descriptions of varying abstraction in communication processes might seem of rather low importance, given that the content of the transmitted message remains clearly positive or negative: regardless whether we state that ‘Aliah was on time for our meeting’ or ‘Aliah was punctual’ we seem to deliver a positive information about Aliah. Yet, although both statements seemingly tell us the same, they use different linguistic categories conveying differing levels of abstraction and thereby decode different information since the different linguistic categories are associated with different psychological functions. Semin and Fiedler (1988) have conducted an analysis on psychological functions of words we use in interpersonal communication from which the Linguistic Category Model (LCM) originated. The LCM basically demonstrates how linguistic abstraction signals the impression an observer has formed of another person, i.e.,
without explicitly saying so, the observer reveals details about the impression he or she has of another person simply by choosing a specific level of language abstraction.

In order to illustrate these consequences of language abstraction and the characteristics of the respective linguistic categories, we extracted two behavioral episodes from the storybook designed for this research project – one positive episode, one negative – and use them as an example for demonstrating the postulations of the LCM.

The LCM constitutes a classification system for words, consists of four linguistic categories that are distinguished by their respective cognitive-psychological functions and implications for person perception and are organized on a continuum of abstraction. These categories are formed by (1) descriptive action verbs, (2) interpretive action verbs, (3) state verbs and (4) adjectives. Our illustration will follow this organizational principle.

This first category of the LCM consists of descriptive action verbs (DAVs). When using DAVs for communication, the speaker solely relies on directly observable courses of action. Since this way of communication forgoes any interpretation, e.g., why a behavior occurs or what relation exists between the actors, and only focuses on the description of the action itself, descriptions of the very same behavior should be almost identical even when delivered by different observers. A further characteristic of DAVs is that they all own an action-specific invariant feature, e.g., the leg movement necessary to kick someone or the facial muscles necessary to produce a smile. Accordingly, all DAVs as well refer to actions with a clear-cut execution, i.e., a notable and apparent beginning as well as ending.

Using DAVs to illustrate the behaviors depicted in Fig. 1, respectively Fig. 2, thus would produce statements like “A is pushing B off the swing. (Fig. 1)” or “A is kissing B on his cheek.” (Fig. 2).
The second category refers to interpretive action verbs (IAVs). Communication in IAVs already enfolds a slightly more abstract level. An IAV is similar time-framed as a DAV, i.e. has a clear beginning and ending, does however not refer to a single specific behavior, yet to a specific class of behaviors. That is, an IAV can code several kinds of specific actions and thus entails a certain level of interpretation. As a consequence, the utilization of IAVs for description of behavior leaves the specific course of action in the dark and provides a positive or negative connotation to events. For example, using an IAV to describe the example behavior in Fig. 1 could produce the statement “A is hurting B” Although this statement gives an evaluative (negative) impression of events, without the given example picture of the scene, a recipient of the statement could not clearly and unambiguously decode which specific action the speaker refers to by hurting. The use of the word “hurting” does not only decode the act of pushing someone but could stand for quite different actions such as spitting, kicking or insulting with
words. The same is true for reporting a positive course of events, e.g., by referring to Fig. 2 with the statement “A is consoling B.” Again, without reference to the example figure which is depicting a kiss on the cheek, the verb consoling could as well stand for stroking someone’s hair, holding someone in the arms or offering some encouraging words.

The utilization of the LCM’s third category, state verbs (SVs), detaches descriptions from a specific action or even actions in general and instead provides a more person-related level of meaning. SVs typically characterize the relation between an actor and the target of the action. Consequently, different from DAVs and IAVs, which refer to temporally clear-cut behaviors, SVs refer to enduring emotional or cognitive states. Descriptions of the example pictures using SVs, thus, could produce the statements “A detests B.” (Fig. 1) or “A likes B.” (Fig. 2).

Referring to foregoing psycholexicological studies (Miller & Johnson-Laird, 1976), Semin and Fielder (1988, 1991) further differentiated SVs from the preceding categories by the easiness with which a word can be used in the progressive form. The use of progressive form is easily done for action verbs like run (“is running”) or IAVs like help (“is helping”) but not applicable for SVs like detest (“is detesting”) or like (“is liking”). A similar way of distinguishing SVs from both previous categories allows the application of the imperative, which is common for action verbs, e.g., “Run!” and interpretive action verbs “Help!”, yet impractical for state verbs, e.g., “Detest her!” or “Like him!”.

The fourth and final word class of the LCM is built by a grammatically common linguistic category, namely adjectives. Adjectives form the only category of the LCM that is already defined by formal grammar rules. Semin and Fiedler (1988) have emphasized that adjectives are used to classify and discriminate persons by associating them with certain attributes. That is, adjectives are purely person-related, they contain personal and characterizing
information while SVs deliver relational information and DAVs, IAVs disclose an action. Referring to our sample behaviors, these could be described by statements like “A is aggressive.” (Fig. 1) or “A is caring.” (Fig. 2) when using adjectives for descriptions.

Recapitulating our “declination” of LCM categories, we can conclude that, although all statements refer to the same illustrations and contain equally true representations of the depicted situation, the statements seemingly paint a somewhat different characterization of events. More specifically, these statements reveal what impression the observer, that is the person giving the statements, has formed about the depicted scenes respectively the acting characters. Thus, these illustrations serve as examples on how linguistic abstraction conveys already formed impressions, since “psychological properties ... systematically vary across the four categories” (Semin & Fiedler, 1988, p. 561). That is, linguistic category systematically vary regarding impressions of (1) enduringness, meaning the more abstract the linguistic category used for description the more we evaluate the presented behavior as enduring, i.e. we estimate the likelihood of behavioral repetition higher; (2) subject and (3) situative informativeness, meaning the more abstract the linguistic category used for description the more diagnostic the information seems about the person and less about the situation, i.e. we assume that the depicted behavior is rooted in characteristics of the person yet not forced by situational circumstances; (4) verifiability and (5) disputability, meaning the more abstract the linguistic category used for descriptions the more difficult is a verification or disputation of its content (Semin and Fiedler, 1988, 1991, 1992).

In other words, the more abstract people talk about somebody the more likely they transport positive or negative evaluation of the person that was subject of the communication. For example, talking about about ‘Ali who slapped someone during a soccer match’ delivers
more information on the context, i.e., what (slap) happened where (soccer match) compared to the statement that “Ali is aggressive” which seems to deliver more information on the subject, i.e. the observer’s evaluation of the subject (as being aggressive). Consequently, with the latter statement it seems far more likely that the observer conveys his impression of Ali having an aggressive personality and such the according expectation of respective aggressive behavior in the future (enduringness). This evaluation of Ali is moreover far more difficult to contradict than an impression that is reported in more concrete. A single event like slapping someone can directly be observed and therefore is more easily verified and more difficult to dispute compared to a general evaluation of a person, like “Ali is aggressive”, for whose confirmation or rejection far more information is needed (verifiability and disputability). That is, describing Ali as aggressive presents a far more stable evaluation of Alis personality than a concrete description would have done. While Semin and Fiedler (1988) reported the psychological function of linguistic categories with respect to the information it reveals on the impression that is formed by the observer, i.e. the person giving the statement in question, it is feasible that these psychological function of linguistic categories affect the receiver of a given statement as well and thereby contribute to form an impression of a person. That is, a person that is told how aggressive Ali is might gain the impression that this is a stable personal attribute, a personality trait and therefore assume that there is a likelihood for occurrence of aggressive acts in the future while reporting a single incident might rather draw more attention to the circumstances of the act, i.e. question what lead to Ali acting in a certain way. Table 2 gives a short recap on all the linguistic categories as suggested by the LCM.

Assuming a causal link between language, more specifically language abstraction, and impression formation the taxonomy of the LCM is a fruitful approach to test whether the
reported affects regarding transmitting already formed impressions might as well extend to impression formation. Whether this is the case and how this might influence attitude formation is to this point however not clear. It will therefore be a main focus of this thesis to test whether systematic variation of linguistic categories in communications about group differently affect impression formation in line with the demonstration of linguistic effects regarding impression representation.

Table 1. Classification criteria for linguistic category as suggested by Semin & Fiedler (1988) and in an extension by Carnaghi et al.(2008).

<table>
<thead>
<tr>
<th>Category</th>
<th>Criteria</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Descriptive Action Verbs</td>
<td>Have a clear defined beginning refer to actions with / end possess a physical invariant feature are neutral</td>
<td>He visits the animal shelter often.</td>
</tr>
<tr>
<td>(DAV)</td>
<td></td>
<td>Further DAV Examples: kick, kiss, run (…)</td>
</tr>
<tr>
<td>Interpretive Action Verbs</td>
<td>refer to actions with clear defined beginning / end refer to a multitude of actions with identical meaning indicate a positive or negative evaluation</td>
<td>He helps animals.</td>
</tr>
<tr>
<td>(IAV)</td>
<td></td>
<td>Further IAV Examples: comfort, protect, attack (…)</td>
</tr>
<tr>
<td>State Action Verbs</td>
<td>refer to general behaviors with a clear beginning / end Refer to emotional consequences of actions</td>
<td>He is excited about animals.</td>
</tr>
<tr>
<td>(SAV)</td>
<td></td>
<td>Further SAV Examples: surprise, amaze, ashamed (…)</td>
</tr>
<tr>
<td>State Verbs (SV)</td>
<td>refer to enduring emotional / cognitive states describes the relation to specific persons / objects</td>
<td>He loves animals.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Further Examples: admire, hate, adore (…)</td>
</tr>
</tbody>
</table>
Research on the LCM has moreover triggered subsequent research connecting language abstraction and linguistic structure to a number of interesting phenomena regarding impression formation. All of these results hint in the direction that abstract rather than concrete language transports the information that there are stable representations of a person, i.e. that acts are rather attributed to personality and / or disposition and that respective representations are not easily changed. Based on the assumption that language abstraction not only represents existing representations but as well shapes impression formation these results could as well be used to further demonstrate how language abstraction influences impression formation.

Fiedler, Semin and Bolten (1989) did, for instance, demonstrate that communicative acts are performed in abstract terms when the speaker is certain about the transmitted information and in concrete terms when the information is still in need of verification. That is, abstract language transports certainty and might therefore be considered a useful source for impression formation.

Moreover, language abstraction has been associated to attribution processes, i.e. the criteria used to determine whether to explain the cause of behavior by external, situational or internal, dispositional inferences (Semin and Fiedler, 1991). For instance, the utilization of concrete terms is related to intentionality (Jones & Davis, 1965), controllability (Weiner, 1985) or distinctiveness (Kelly, 1967) and, thus, typically used in situational attribution processes. In contrast, rather abstract expressions are associated with criteria in favor of dispositional
inferences such as (high) consensus (Kelly, 1967) and (high) stability (Weiner 1985). Rather abstract language, especially adjectives, are frequently used for interpersonal descriptions. This is caused by the prevailing tendency to infer traits and dispositions even when the baseline descriptions are given in action verbs (Semin, Fiedler & Bolten, 1989).

That is, abstract language might allow for more dispositional inferences, thereby more easily leading to personality-based evaluations in impression formation.

Language use can also influence the confirmability of a statement. Rothbart and Park (1986) demonstrated how using adjectives to report negative qualities of a person creates a persisting negative impression of a person because negative traits are easy to confirm and hard to disconfirm. This pattern is however reversed, when using adjectives to describe positive qualities of a person. Yet, it is not only the valence of adjectives that affect the confirmability of a person’s characterization. Confirmability can also be influenced by, firstly, the trait dimension indicated by an adjective, e.g. whether an adjective indicates warmth or competence (Tausch, Kenworthy & Hewstone, 2007), and secondly, by the receiver’s individual conceptions about the dispositional character of personality traits (Dweck, Hong & Chiu, 1993).

That is, abstract language might lead to more stable and enduring impressions – yet there might be linguistic nuances within linguistic categories that affect impression formation beyond the systematic variation that could be derived from the LCM.

The quintessence from all this empirical evidence is, that (a) the LCM provides a taxonomy of interpersonal words that systematically vary regarding the psychological functions associated with these categories and demonstrates how language ‘reveals’ existing representations and (b) additional research supports the assumption that abstract language represents more stable, dispositional and less disputable impression. Even though there is yet no
empirical research on this specific topic, it might be possible that these effects are not limited to impression representation but extend to impression formation as outlined above. Providing empirical evidence regarding this hypothesized link between language and impression formation is the focus of this thesis – specifically regarding the role of language concerning impression formation in intergroup contexts.

**How Words Represents Biases in Inner Worlds: The Linguistic Intergroup Bias**

Language use in intergroup contexts provides an interesting field of application for the LCM. Using the taxonomy of the LCM Maas and colleagues (1989) used variations of language abstraction in intergroup situations to determine whether intergroup evaluations might be derived from language use. In their initial experiments they used a very specific intergroup context, namely among supporters of a traditional Italian horseracing competition. These races, the so-called *palios*, are deeply rooted in Italian history and still are an important part of regional culture. During a palio members of several neighborhoods, named *contradas*, annually compete against each other.

In advance to one of these *palios* Maass and colleagues (1989) investigated how members of the involved *contradas* tended to describe behavior of their own group and the behavior of other *contradas*, respectively. For that purpose, Maass et al. (1989) developed a research design that quickly became standard procedure for investigations of linguistic biases. Basically, participants were presented with several cartoons depicting members of their own group or members of a competing *contrada* engaging in positive, socially desirable and in negative, socially undesirable behaviors. An example cartoon is given in Figure 3.
Each cartoon came with a selection of four additional descriptions of the depicted behavior, which at the same time provided the response alternatives for participants. While all descriptions referred to the action presented in the cartoon, the statements varied in abstraction. For example, the situation depicted in the cartoon below (Fig. 1) was described by a concrete statement like “The actor gives the horse an injection” or an abstract statement like “The actor cheats”. Participants were asked to choose the statement that described the cartoon best.

Figure 3. Example cartoon used by Maass et al. (1989), extracted from Maass (1999), p. 82

<table>
<thead>
<tr>
<th>Valence of Behavior</th>
<th>Observed Behavior</th>
<th>Agent of Behavior: In-Group Member</th>
<th>Agent of Behavior: Out-Group Member</th>
</tr>
</thead>
<tbody>
<tr>
<td>Positive</td>
<td>Hospitality</td>
<td>He was cordial / hospitable.</td>
<td>He offered food and drinks.</td>
</tr>
<tr>
<td>Positive</td>
<td>Helpfulness</td>
<td>She was very friendly / helpful.</td>
<td>She gave us directions when we were lost.</td>
</tr>
<tr>
<td>Negative</td>
<td>Aggression</td>
<td>He slapped the referee in the face.</td>
<td>He was hostile / aggressive.</td>
</tr>
<tr>
<td>Negative</td>
<td>Discourtesy</td>
<td>She didn’t greet me when walking by.</td>
<td>She was disrespectful / rude.</td>
</tr>
</tbody>
</table>

When analyzing participants’ responses, Maass et al. (1989) observed that chose more abstract descriptions for positive behaviors of their own contrada and negative behavior of the
competing *contradas*, while they chose rather concrete descriptions to refer to negative behavior of their own *contrada* and positive behaviors of the competing *contrada*. This specific linguistic distinction between groups was named “Linguistic Intergroup Bias (LIB)”. In general terms, the LIB is a specific pattern of language use in the intergroup contexts with abstract descriptions being used for positive ingroup and negative outgroup behavior respectively concrete descriptions being used for negative ingroup and positive outgroup behavior.

Although the LIB was originally observed within a very specific intergroup context for a culture-specific tradition, it has proven to be a more general phenomenon. Over the last decades linguistic biases have been demonstrated for a variety of intergroup contexts such as gender (Wigboldus, Semin & Spears, 2000), political parties (Rubini & Semin, 1994), local affiliation (Maass, Milese, Zabbini & Stahlberg, 1995), sport teams (Franco & Maass, 1996; Tanabe & Oka, 1996), racial or ethnic groups (von Hippel, Sekaquaptewa & Vargas, 1997) and in different languages, such as Italian (e.g., Maass et al. 1989), German (e.g., Schoel, Roessel; Jacobsen & Stahlberg, 2014), Dutch (e.g., Werkmann, Wigboldus & Semin, 1999), English (e.g., Karpinski & von Hippel, 1996), Chinese (e.g., Ng & Chan, 1996), and Japanese (e.g., Tanabe & Oka, 2001).

While a major part of these studies on linguistic biases has been conducted in laboratories and therefore applied different variations of the forced choice paradigm typical for research on the LIB, there is also profound evidence for linguistic biases in spontaneous language use (e.g., Rubini, Graziani & Moscatelli, 2009) and in real-life communication processes, especially media coverage (e.g., Dragojevic, Sink & Mastro, 2016; Geschke, Sassenberg, Ruhrmann & Sommer, 2010; for influences of television-media reports on linguistic biases in viewers see Gorham, 2006).
With its pervasion in a multitude of languages, cultures and contexts, the LIB therefore appears to be a strikingly general and stable phenomenon. Although research on the LIB has provided compelling evidence on how language demonstrates biased intergroup attitudes and it is therefore likely that language plays a significant role in intergroup relations they don’t allow for secured derivations on language functions in intergroup contexts, specifically the question whether language causally affects attitude formation.

**Cognition or Motivation – What drives the Linguistic Intergroup Bias?**

Examining mechanisms driving the LIB enables not only a more comprehensive understanding of the effect itself but provides a first approximation to its function in the intergroup context, too. That might as well enable a better understanding of its potential function in the developmental context regarding the evaluation of groups respectively the formation of intergroup attitudes. Over the years two mechanisms were considered to impel linguistic biases in intergroup related speech: motivation on the one hand, cognition on the other hand.

Assumptions that motivation is causing the LIB are primarily based on SIT (Tajfel & Turner, 1978, 1979). SIT suggests that we thrive for positive images of our self and in partial derive this positivity from positive evaluations of one’s social identity, thus people strive to maintain memberships in groups that are distinctly positive from other groups. Intergroup comparisons do however not always result in positive distinction. SIT identified several strategies to create a positive self-image despite the negativity of group membership. These are known as (1) individual mobility, (2) social creativity and (3) social competition. The latter (3) basically describes a process in which groups at a disadvantage will enter into direct competition for rare resources with higher status groups.
Strategies of (1) individual mobility are dissociation and disidentification, both applicable only when group boundaries are found to be permeable, i.e., the process of exiting one and joining another group can be done with ease, i.e. altering political views as opposed to altering gender (see e.g., Smith & Mackie, 2007, p. 219). Dissociation is a process of physical distancing from the group, i.e., weight loss in order to leave the group “obese people”, or concealment of group membership, i.e., concealing religious affiliation by foregoing to wear typical religious symbols or clothing. Disidentification is a form of psychological distancing, e.g., shifting identification from a negative evaluated group to another more positive evaluated group (Mussweiler, Gabriel & Bodenhausen, 2000), emphasizing the heterogeneity within the group to allow disidentification with negative group features (Doosje, Ellemers & Spears 1995; Ellemers, Spears & Doosje, 2002) or alienation by open criticism, the so-called black sheep effect (Branscombe, Wann, Noel, & Coleman, 1993; Lewis & Sherman 2010; Marques, Yzerbyt & Leyens, 1998.).

Forms of (2) social creativity generally focus on renewing or improving the evaluation of the group. This can be done by finding new dimensions of comparison, more precisely, dimensions in which the own-group outperforms others, e.g., an obese person comparing himself on the dimension of humor instead of sportiness, completely exchanging the comparison groups, e.g., comparing the soccer team to the less successful football team instead of the more successful baseball team, or redefinition of groups’ value or the value of significant group-attributes. The movement “Black is beautiful” is a well-known example of the reevaluation of a highly-discriminated group by reversing the meaning attached to the essential distinction criterion of the group (see Smith & Mackie, p.221).

The alternative approach forgoes motivation and explains the LIB with a cognitive
mechanism. Similar to the above-described mechanism, the cognitive approach is related to another form of intergroup bias. Research on memory performance in intergroup contexts has previously illustrated that people tend to have biased expectations for intergroup related behaviors (Howard & Rothbart, 1980). People expect more positive and less negative behaviors from the ingroup compared to the outgroup. The difference in language abstraction, as demonstrated by the LIB, could thus be related to differential behavior expectations. When it is assumed that expected behaviors originate from stable dispositions, the use of abstract language to describe expected behaviors seems natural within the logic of the LCM. At the same time unexpected, sudden or surprising actions are probably best described by the use of concrete terms, i.e. terms that allow for situational rather than dispositional attribution of the behavior (Beukeboom, 2014).

Although both approaches are seemingly different, both often yield aligned predictions. For example, given the intention to describe positive ingroup behavior, the motivational approach would predict abstract language (“helpful”, “supportive”) aiming to create the most favorable representation of the ingroup. The cognitive approach would argue for the use of abstract language as well, as long as the behavior is expected. As the prevailing tendency is to expect rather positive behavior from ingroup members (“helpfulness” or “support of one another”), the cognitive approach would likewise predict the use of abstract language when describing helpful conduct of ingroup members. However, both approaches diverge when expectations deviate from the typical scheme, i.e. expecting negative ingroup and positive outgroup behavior. This can be the case when stereotypes about outgroups contain positive characteristics, e.g. people of Turkish descent are believed to be warm and generous hosts. When relying on the motivational approach, the observer would report positive outgroup behavior on a
rather concrete level. Thus, an example sentence might be: “This (turkish) woman offered tea to the guests.” According to the cognitive approach, however, expectations about behavior of Turkish women are decisive for choice of language abstraction. With the expected caring behavior towards guests; thus, the observer would rather give his statement on a more abstract level. An exemplary statement would be: “This woman was very hospitable”.

Attempting to identify which of these approaches is the driving force of the LIB, researchers focused on circumstances under which the motivational and cognitional approach yield different predictions. Experiments designed for this purpose subsequently provided evidence that untypical, stereotypically incongruent behavior was typically described in concrete terms, while abstract language was used to describe stereotype-congruent behavior – regardless of the valence of behavior (Maass et al., 1995; Rubini & Semin, 1994). That is, contrary to postulations of the motivational approach, it is not the behavioral valence but rather behavioral predictability that critically influences the level of language abstraction in communication. Even more conclusive in this regard are results demonstrating the use of abstract language for expected behaviors in absence of a need for ingroup protection motives, i.e. in circumstances in which social belonging or group status remain unmentioned. This holds to be true when the subject of communication is an individual of personal relevance about whom participants already hold expectations, i.e. friends or enemies, but as well when the subject of the communication is unknown or fictitious, i.e. when the subject is an individual about whom expectations are experimentally induced (Maass et al., 1995, Exp. 2 and Exp. 3).

The apparent importance of expectancies as a driving force for linguistic biases has led researchers to refer to expectancy induced linguistic biases with an own term, namely the Linguistic Expectancy Bias (LEB) (Maass, 1999; Wigboldus, Semin & Spears, 2000). Further
investigations on the LEB revealed that, besides expectations, it is moreover the social context of a communication process that is important for the production of linguistic biases.

For example, when reporting stereotypically consistent gender behavior, the expectancy bias emerges only in conversations between members of the opposite gender, i.e. male-female-conversations, but is absent in communication processes between same gender members, i.e. male-male or female-female conversation pairings (Wigboldus, Spears & Semin, 2005). Wigboldus and colleagues (2005) explained this context dependency of the LEB by relying on stereotype activation as a prerequisite for the LEB. Hence, it is assumed that social context determines which categories are salient in a communication process. Accordingly, activated stereotypes, consequently generate corresponding stereotypic expectancies, which finally results in an LEB. Wenneker, Wigboldus and Spears (2005) moreover observed that people tend to use biased language only when they heard about the social belonging or group membership of a communication target before the actual information was given. In contrast, when people first received information about someone and learned about his or her group membership only afterwards, there was no bias in language visible, i.e the process of information encoding seemed to be the cause of the bias.

Bringing together the research of Wenneker et al. (2005) and Wigboldus et al. (2005) shows that the salience of group membership is essential for the production of the LIB. Firstly, linguistic biases seem to occur only in intergroup-contexts, i.e. when the group-membership of at least one actor differs from the others. And, secondly, only when group-membership is activated before information about the respective person is processed, a LEB emerges.

Following the argumentation of Maass (1999) it is to assume that expectancies are sufficient to produce linguistic biases, i.e. expectancies can be considered the default force
advancing linguistic biases. That is, as soon as parents, caregivers or other adults have stereotypic representations of specific groups and the intergroup context is salient, linguistic biases will likely occur and thereby might as well transmit biases in expectation of social groups on children and that these biases might be stronger when according motivation is given. That the LIB itself is susceptible to motivation, specifically to the need for ingroup protection, i.e. biases being larger in the presence of threat, has been demonstrated in laboratory settings (Maass et al. 1994, Exp. 3) as well as in real life communication processes (Maass, Ceccarelli & Rudin, 1996). Through the specific pattern of creating ingroup positivity and outgroup negativity the LIB maximizes the evaluative difference between both groups. Following the idea of language causally influencing group representations it might be feasible that motivation does not only influence the scope of the LIB but that it also contributes to the formation of positive social identity, which in return also bolsters individual self-esteem (Maass et al., 1989; Maass & Arcuri, 1992; Maass, 1999). Thus, the LIB might function similar to other strategies of social creativity maximizing evaluative differences between in- and outgroup resulting in a maximal positive outcome for the ingroup in verbally made comparisons. It seems at least possible that parents transmit group evaluations by language in the attempt to enhance social identity, specifically under circumstances in which said identity is threatened.

That is, neither cognition nor motivation seem exclusively responsible in driving the LIB yet rather are complementary forces. Both principles influence the occurrence of linguistic biases and thereby might as well influence the formation of group representation. On a baseline level, language could transmit learnt expectations about social groups and thereby shape learning of novel social groups in childhood (cognition) and under specific circumstances biases might be increased and learning of group evaluations thereby facilitated (motivation).
Linguistic Biases and Stereotypic Beliefs: A Hen and Egg Problem

Language is the key tool of human communication and language biases are seemingly a vital part of our communication processes. Considering how expectations, communication goals and ingroup protectiveness shape language use, it is a natural conclusion that “the linguistic bias phenomenon (…) has important implications for the communication of (social) information in general.” (Wigboldus et al., 2000, p. 16). So far there has been none empirical evidence regarding the effect of linguistic bias on attitude formation, yet it has often been assumed that there is a causal link to some extent.

The probably most important implication of linguistic biases is its association with stereotype perpetuation. Several researches have assumed that communication of stereotypic social beliefs in abstract language contributes crucially to the stabilization of stereotypic belief-systems and likewise aggravates their falsifiability (Maass et al. 1989, Maass & Arcuri, 1992; Maass, Arcuri & Suitner, 2014; Schoel, Roessel, Jacobsen & Stahlberg, 2014). It is easily understandable that, if communication about transgression of outgroup members is reported in abstract language, e.g. talking about outgroup members as dishonest, aggressive, lazy, corrupt and ruthless, it can foster a negative stereotypic impression of the outgroup. Consequently, it has been assumed that the LIB as a forthgoing communication pattern in intergroup contexts constitutes an important source of interpersonal transmission of (stereotypic) beliefs and social attitudes (Geschke et al., 2010; Maass & Arcuri, 1992; Maass, 1999; Wigboldus et al., 2000). It has thus been concluded that linguistic biases are not only associated to (biased) social attitudes, but might be accountable for the stabilization, perpetuation and transmission of (biased) social attitudes.
Assuming that linguistic biases are indeed related to stereotype transmission and perpetuation, it raises the question on the nature of this association. More precisely, it raises the question whether biased language is the consequence of biased intergroup attitudes or whether biased language might be the cause of biased intergroup attitudes. Early research on the LIB clearly defined this bias as a result of pre-existing intergroup bias (Maass et al., 1989) and thereby conceptualized the LIB as a reflection of previously established intergroup attitudes (Maass et al., 2014). However, some of the conclusions made in the later course of the research process seem to challenge this view and postulate that language might assert socialization influences on attitude formation. Werkmann and colleagues (1999) have, for example, proclaimed that linguistic biases might be associated with the formation of intergroup attitudes; Maass herself (1999) considered that linguistic biases might contribute to the creation of social representation and Schoel and colleagues (2014, p. 274) considered linguistic biases as “critical factors in stereotype development”.

Notwithstanding this compelling conclusion, up to this point there is only scarce evidence on a socialization function of language in intergroup contexts – neither in the social psychological literature nor in the developmental literature. In general, there is very little research explicitly targeting the question whether linguistic variations influence intergroup attitude formation.

One of the few studies explicitly investigating the effects of language bias on the recipient of a message was conducted by Wigboldus and colleagues (2000), who established an empirical link between language biases in interpersonal conversation and subsequent biases in impression formation. In the respective study, Wigboldus et al. (2000) have demonstrated that stereotypic behavior is communicated more abstractly and, even more important with the regard
to the hypothesis of language causally influencing impression and attitude formation, that participants draw more dispositional inferences when having listened to behavioral descriptions in abstract language. This specific relation between abstract communication and dispositional inferences, applied for communications emphasizing the social category of gender (Exp. 1 and Exp. 3) and in parts for nationality (Exp. 2, however, this experiment involved only participants of one nationality, i.e. Dutch, and these participants drew more dispositional inferences when listening to abstract language only for reports about the Flemish outgroup, yet not more dispositional inferences were drawn when listening to abstract messages about the Dutch in-group). Regarding the specific mode of communication, the reported relation between language and inferences, was observed for behavioral descriptions spontaneously made up in own words (Exp. 1), for given descriptions that had to be spontaneously retold in own words (Exp. 2) and for experimentally generated, controlled behavioral descriptions. Wigboldus and colleagues (2000) thereby successfully demonstrated a relation between language abstraction and impression formation. Although this relation has been discussed in previous research already, with, for example, Semin and Fiedler (1988) demonstrating systematic influences of language abstraction on person-related inferences, Wigboldus and colleagues (2000) are possibly the first to establish a direct association between variations in spontaneous language use and impression formation.

While the presented results might be a convincing indication, how social attitudes can be transmitted in interpersonal communication, these do not provide evidence for a causal influence of language on attitude formation. It is, especially, the specific design that is not suitable to allow conclusion on causal effects of language. Firstly, the studies worked with gender and nationality, both pre-existing social categories. We, therefore, cannot decide whether language initially
caused inferences about these categories or whether previous knowledge on these categories has determined the inferences made. Secondly, all experiments assessed how impression formation might be influenced with respect to specific targets. For example, in Experiment 1 participants were asked to think about a friend and communicated background information and behavioral episodes about this specific friend. Accordingly, the dependent variables estimated the likelihood that this friend repeats a certain behavior respectively the likelihood with which either situational or dispositional variables determined the behavior of this specific friend. Similarly, Experiment 2 investigated participants’ responses to a specific Dutch or Flemish target and Experiment 3 investigated responses inter alia to a specific female (Sandra). Conclusively, the design of the experiments was cut out to assess the effect of language bias on the evaluation of a single, individual target. It is not clear how these can be transferred to linguistic influences on actual evaluations of groups, let alone linguistic influences on the formation of social representations of groups.

In order to determine whether intergroup attitude formation can be causally affected by linguistic biases, it seems advisable to alter experimental specifics. Consequently, research needs to focus on the effects of language abstraction on the evaluation of social groups and more specifically, concentrate on social groups for which yet no representations have been established. From research in the area of developmental psychology, we know that children acquire important social information at an early developmental stage and likewise build attitudes towards social categories early. More precisely, the analysis of Raabe and Beelmann (2011) indicated a peak of prejudice between the age of five and seven. Categorization of social groups and formation of intergroup attitudes is likely to happen before. That is, principles, mechanisms and functions contributing to formation of intergroup attitudes possibly to come to effect already in early
childhood. Although Bigler and Liben have suggested language as a relevant factor in attitude formation, there is yet no information on possible effects of linguistic biases. We know, however, from research of Gelman and Heymann (1999) that children seem to be susceptible to linguistic nuances in language by the age of five to seven. Children, for example, drew different inferences about new characters depending on the linguistic level the information was presented to them. Combining these results from developmental and social psychology as well as psycholinguistics it seems recommendable to analyze potential influence of language abstraction on the acquisition of social information in early childhood. We therefore conduct our experiments mainly with kindergarten children, assuming that if language has a causal effect on attitude formation this effect should be visible during the phase in which categorization and initial group evaluation takes place respectively begins.

**Linguistic Intergroup Bias from a Developmental Perspective**

Researching evidence in favor of the assumption that children’s acquisition of social representations and attitude is shaped by language, it has to be stated, that social psychological as well as developmental psychological literature has provided rather scare evidence for the socialization function of language. There is altogether only a meager amount of research that explicitly targets the question whether linguistic variations influence intergroup attitude formation in general or in children and adolescents especially (see as well Schoel et al., 2014).

The few available studies have primarily focused on linguistic biases in communication among adolescents. Salès-Wuillemin, Masse, Urdapiletta, Pullin, Kohler and Gueraud (2014) as well as Schoel and colleagues (2014) have investigated language biases in adolescents in an ethnic intergroup setting, the first in France and the latter in Germany. For the German sample,
Schoel et al. (2014) demonstrated that German and Turkish adolescents communicate stereotype-congruent information more abstractly, confirming a general LEB. However, a LIB was shown only under circumstances motivating ingroup protection, i.e. when classroom composition produced a foreign majority, when students perceived a high learning disadvantage or when interethnic conflicts were frequent. And even more importantly, the experiment was designed to demonstrate how children use linguistic biases as a reflection of or to signal their intergroup attitudes. It does not allow to draw conclusion how linguistic biases might affect attitude formation. While Schoel et al. (2014) focused their research on adolescents, i.e. participants whose age ranged between thirteen and 20 years, Salès-Wuillemin and colleagues (2014) investigated language bias with younger children with an age range from seven to eleven years. Different from Schoel and colleagues (2014), Salès-Wuillemin et al. (2014) found language bias only regarding ingroup favoritism. That is, positive ingroup behaviors were described rather abstractly while negative ingroup behaviors were described rather concretely. However, language use for descriptions of outgroup behaviors did not seem to be biased – neither for majority nor for minority children. An additionally employed linguistic marker, the frequency of mentioning an ingroup target when describing a positive behavioral episode, similarly demonstrated ingroup favoritism. Salès-Wuillemin et al. (2014) offered two explanations for the accordance of majority and minority children’s response patterns – both focusing on a rather untypical self-perception of minority children. The authors assumed that the minority group, black children of sub-Saharan African descent, either perceived their minority status as unstable or interchangeable. Another explanation was that, due to relatively high rates of minority children in the participating schools, minority children did not perceive themselves as part of a minority – thus yielding alike responses for both groups. Even though Salès-Wuillemin et al.
DEVELOPMENTAL PERSPECTIVE ON LIB

(2014) demonstrate that linguistic biases are found in younger children as well, the results again are meaningful for the LIB as a form of attitude reflection, yet allow no conclusion on language effects regarding attitude formation.

Diverging from an ethnic intergroup context, Werkmann and colleagues (1999, Exp. 1) demonstrated that children and adolescents from eight to 19 years, produced a LIB when describing imaginary positive and negative behaviors of their best friend or their worst enemy. Even more important is, that in this study the LIB seemed to increase with age (Werkmann et al., 1999, Exp. 1), however the same limitation applies to the study of Werkmann and colleagues (1999) that has been described for Schoel et al. (2014) and Salès-Wuillemin et al. (2014): results demonstrate linguistic biases production in childhood but are not informative regarding linguistic biases influencing attitude formation.

The combination of these three studies substantiates that language bias in intergroup contexts appears across cultures early in development and partially can provide evidence that language biases intensify with age. Yet, none of these studies has directly linked language to the formation social attitudes. It seems that children use language as a reflection of already formed evaluations of groups but the design of the presented research allows no conclusion on whether linguistic biases used in social communication actually influence children’s evaluations of novel group. So even though it has been assumed that linguistic biases contribute to cross-generational stereotype formation and this assumption seems convincing regarding the research demonstrated in the LIB and its potential fit to the theoretical frameworks presented above, there is yet no empirical research supporting the idea that linguistic biases causally influence attitude formation.

To our knowledge there is only one published study, published two decades ago, in which the effect of language abstraction on children’s inferences was directly investigated (Werkmann
et al., 1999, Exp. 2). In that study, Werkmann and colleagues (1999; Experiment 2) observed that children as young as five years evaluated repetition likelihood for trait-related behaviors, i.e. being funny or being talkative, higher, when previous target behaviors were reported in abstract compared to concrete terms. Also, children as young as seven years rather attributed behavior to personal instead of situational factors when respective behavior was described in abstract compared to concrete terms.

By demonstrating that children draw inferences based on the abstraction of behavioral descriptions, this experiment delivers first proof of an early susceptibility to language abstraction regarding impression formation. Providing the knowledge that children differentially react to language abstraction, is a first step towards determining whether language use contributes to the formation of group representations and intergroup attitudes during childhood.

Similar to the research design used by Wigboldus et al. (2000), Werkmann and colleagues (1999) investigated the inferential potential of language with respect to the actions of an individual target and accordingly used dependent variables assessing future behavior and personality aspects of this very target. Hence, the inferential potential of language abstraction for children should be interpreted regarding the study design, i.e. in regard to impression formation in individuals—not groups. Nevertheless, the results of this study deliver valuable insights on how linguistic biases might shape our representation of the social world. We do however not know whether the inferential potential demonstrated for individual targets similarly occurs for groups, i.e. whether formation of group evaluation is equally affected by linguistic biases.

Valuable conclusions might moreover be derived from studies that focus on the question how verbal cues affect the acquisition of social representations and categories. Although, these studies do not specifically investigate the effect of language abstraction, they examine how
language affects a process that is an important part of attitude formation.

For example, Dunham, Baron and Carey (2011) demonstrated that randomly assigning young children to minimal groups was sufficient to induce intergroup bias in children. This bias was subsequently visible in attitudes, in partial also in measures of resource allocation (in Exp. 1 but not in Exp. 2) and behavioral attribution (in Exp. 2 but not in Exp. 1) as well as play preference (Exp. 3). Interestingly, the strength of this effect was dependent on the type of verbal cue used to assign children to minimal groups. Effects seem to be stronger when minimal groups were sorted by using a noun-phrase, i.e. “the orange/green/red group”, compared to experimental conditions in which minimal groups were sorted by descriptive phrases, i.e. “wearing an orange/red/green shirt”, that were presented on a more concrete level. Even though the taxonomy of the LCM does not originally include nouns, subsequent research has investigated nouns as a potential extension of the model. Carnaghi, Maass, Gresta and Arcuri (2008) basically demonstrated for several types of inferences regarding person perception, that the effect of nouns exceeds the effect of adjectives and nouns therefore seem to be the linguistic category with the highest informativeness. The research of Carnaghi et al. (2008) is described in more detail in the introduction of Experiment 3, with which we investigate the effect of noun use regarding group evaluation and attitude formation. Interpreting the results of Dunham et al. (2011) with regard to research of Maass et al. (1989, 1999) and Carnaghi et al. (2008) the pattern of the presented results seems to resemble the prediction that is made concerning the influence of the LIB on attitude formation, since abstract verbal cues (noun-phrase) seem to assert stronger influence on evaluation than concrete verbal cues (descriptive phases). This might be an important cue regarding causality in the influence of group (evaluation) and intergroup attitude formation, however the linguistic nuances in the experiment of Dunham et al. (2011) were only used for
distinction of groups (who is who), whereas we are more interested in the use of biased language regarding behavioral descriptions of the targets.

Even more intriguing is a series of studies conducted by Baron, Dunham, Banaji and Carey (2014) analyzing how visual and verbal cues influence the acquisition of social categories. A central outcome of this research is that the utilization of only verbal labels is sufficient to produce the representation of a social group in children at the age of four years while the utilization of a visual marker only is not. In contrast, older children, about seven years old, built social representations irrespectively of the type of label used; yet children’s inferences about groups were stronger when groups were previously marked with a noun label.

Although these results do not allow direct conclusions about the role of language on attitude formation, they demonstrate the importance of language for social category acquisition that in turn is an essential prerequisite of attitude formation. At the same time, these results challenge central assumptions of SCDT, namely the importance of visual cues for the development of stereotypes and prejudice. Accordingly, Baron and colleagues (2014) conclude that during childhood visual differentiation is more important for identification of individuals rather than for the determination of categories and, thus, probably is of minor importance to stereotype development.

Building on these results, Baron and Dunham (2015) investigated the role of labeling in the formation of ingroup bias. Children showed no signs of ingroup bias when they were assigned to novel social groups that were not verbally labeled, that is children seem to rely on verbally presented group information at least to some extent when forming group evaluations respectively developing ingroup bias. The results are however limited to the use of labels for the distinction of social groups, thereby indicating potential importance of language regarding group
information, yet not informative regarding linguistic biases in the description of ingroup and outgroup behavior.

These studies provide evidence that differential language use at least to some extent can influence children’s impression formation respectively influence whether and how children acquire mental representations of groups or develop ingroup bias. With the following series of experiments, we aim to test the assumption that whether the effect of language goes beyond above cited research and extends to causal affecting children’s acquisition of intergroup attitudes as well.

**OUTLINE OF THE PRESENT RESEARCH PROJECT**

The Linguistic Intergroup Bias describes a subtle but stable pattern of biased language use when reporting in- and outgroup behavior. It has been assumed that this phenomenon plays a subtle but crucial role for stereotype formation (e.g., Maass, 1999; Werkmann, et al. 1998). Relying on theories about stereotype formation, it is indeed conceivable that language abstraction links social categories to distinct attributions and by that fosters either stable positive or negative evaluations of social groups.

Although there is first evidence that language can shape impression formation early in childhood (Werkmann et al., 1998), to this point, there is no research investigating whether and to what extent language abstraction influences the formation of intergroup attitudes neither in adults nor in children.

Therefore, this thesis will focus on the question if, and if so, under which conditions varying levels of language abstraction influence the acquisition of evaluative representations of novel social groups in young children. Results of this research will shed light on developmental
processes in the acquisition of intergroup attitudes, and thus can be a valuable contribution to theories of stereotype formation by linking social and developmental psychology. We designed our experiments to test the hypothesis that language abstraction affects children’s acquisition of novel social groups in a way that (1) the use of abstract language would produce more distinct evaluations of groups with clear valence of behavior, e.g. groups depicted only with positive behavior would be evaluated more positively when behavioral descriptions were abstract rather than concrete and (2) the use of abstract vs. concrete language would result in rather positive or negative group evaluations even though there was no clear valence of group behavior, e.g. groups were introduced with an equal amount of positive and negative behaviors, yet would be evaluated more positively when the presented positive behavioral sequences were described abstractly, while the presented negative behavioral sequences were described concretely.

All experiments introduced children to two novel social groups using age-appropriate pictured storybooks in which we systematically varied the abstraction of the accompanying narratives.

In the first experiment, the two novel groups were clearly and unambiguously linked to either only positive or only negative behavior. We manipulated the level of abstraction of the behavior descriptions between participants. With this experiment we intended to examine whether children acquired representations of novel groups based on behavioral descriptions and if language abstraction influenced the strength of acquired representations.

In the second experiment, we presented children with novel groups that displayed the same amount of positive and negative behaviors. However, we manipulated the abstractness of descriptions of the behaviors in order to test whether this would lead to different evaluations of these groups based on the abstraction of behavioral description. Applying the language pattern of
the LIB, we therefore presented positive behavior of Group 1 in abstract terms while we presented negative behavior of Group 1 only in concrete terms. We reversed this pattern for behavioral descriptions of Group 2. With this experiment we aimed to verify whether subtle manipulation of language abstraction is sufficient to create distinct evaluative representations of groups. We further intended to examine whether higher levels of abstraction in behavior description generate an even greater distinction of group evaluations.

In the third experiment, we again presented two groups displaying the same amount of positive and negative behaviors. Differing from Experiment 2 we altered however the context in which behaviors were depicted. While groups in Experiments 1 and 2 were always depicted in an intragroup context, in Experiment 3 all behaviors will be executed in an intergroup context. Thereby, the design of the third experiment was aligned more closely to the pattern of the LIB for which the intergroup context is a constituting feature. The purpose of this experiment was to test whether subtle linguistic differences are sufficient to create biased representations of groups in an intergroup context when group behaviors are equally positive and negative and moreover if higher levels of abstraction generate more distinct representation of groups. We conduct this experiment not only with children (Experiment 3a) but use an adult sample (Experiment 3b) as a form of control group to investigate whether language ability or cognitive development might affect the results.
Linguistic Manipulation of Group Perception for positive or negative novel groups

(Experiment 1)\(^1\)

Aim of this first experiment was to investigate (1) whether children quickly acquire evaluative representations of novel groups and (2) whether the linguistic level with which groups were presented influence the acquisition of group representations.

For that purpose, we designed a storybook introducing two novel social groups. One of these groups was depicted only with positive behavior while the other group was depicted only with negative behavior. We introduced novel and fictitious instead of real social groups to reduce the likelihood of tapping previously established evaluations of group. We examined how differences in the abstraction of the narratives used to introduce group behavior affect children’s perceptions and evaluations of these groups. We induced these subtle linguistic differences in storybook narratives by relying on the LCM (Semin & Fiedler, 1988). That is, we presented groups with behavioral descriptions that either employed only DAVs or only IAVs or only adjectives respectively, resulting in three experimental conditions. We expected this experiment to provide first evidence on whether children are susceptible to linguistic markers when learning about novel social groups.

Method

Materials.

*Storybook.* For the purpose of this experiment, we created a storybook presenting two novel social groups, namely two fictional kindergarten groups – the Lupis and the Nifos. These

\(^1\) We preregistered our experiment with the Open Science Framework, the registration is retrievable under https://osf.io/m35qz/.
labels are a German adaptation of the labels Lups and the Nifs, which were used in many previous research on the acquisition of social categories in childhood by Baron and colleagues (2014) as well as Baron and Dunham (2015). With our design we built on results of Baron and colleagues (2014) who demonstrated that the use of noun labels is in general sufficient to initiate a categorization process, yet the utilization of visual cues can reinforce categorization for children. We therefore decided to signal group membership by label and an additional visual cue: orange/reddish versus blue shirts.

Both groups were introduced to participants on the first page of the storybook. Participants were presented with group pictures of the Lupis as well as the Nifos and the introductory explanation referred to the group pictures by using the assigned group label and explicitly pointing out the clothing color for each group.

This introductory page was followed by 10 short stories, containing alternating behavioral episodes of either the Lupis or the Nifos. All behavioral episodes took place in an intragroup context and members of the two groups were never displayed interacting with one another. All behavioral episodes were presented on a double-page spread with one picture sized 13 x 16 cm on each page. The page on the left consistently was used to introduce children to the situational context of the behavioral episode, whereas picture and narrative on the right side of the double page revealed the target behavior.

The narrative introducing the situational context was thereby always and only phrased in DAVs and was held identical for all conditions. However, the descriptions of the target behavior varied between conditions in that they either contained a DAV, an IAVS or a person-describing adjective, i.e. we had three experimental conditions. One behavioral episode consisted, for instance, of a picture showing children playing in a sandpit with one child having a whole lot of
marbles, and a second picture showing the child handing some of his marbles to the others. The accompanying verbal description was the same for the first page, introducing the context, i.e. in this case saying “There are other children in the sandpit, and a Lupi, too. The children want to play with the marbles in the sand. The Lupi has a big net holding all the marbles.” Descriptions of the second picture varied depending on experimental conditions, presenting the description of the following scene in varying levels of abstraction, i.e. either stating “Look at the Lupi. The Lupi gives every child exactly three marbles.” (DAV) or “Look at the Lupi. He is sharing the marbles with the other children.” (IAV) or “Look at the Lupi. The Lupi is really generous.” (ADJ). An example page of the storybook with the described scene is given in Figure 4.

Figure 4. Example episode from the storybook in Experiment 1.

All behavioral episodes that were included in the storybook were chosen with relevance to the
horizon of experience for children, i.e. situation that are familiar and relevant for children in their everyday life and that should be within the regular range of experience of kindergarten children. Positive narratives, for example, depicted a child consoling someone or a child sharing something, negative behaviors, for example, depicted a child mocking or excluding someone. An example of our materials including pictorial and verbal stimuli is contained in the appendix. It is important to note that the members of each group consistently engaged in only positive or only negative behaviors, thus creating perfect contingency between group membership and behavioral desirability.

Participants.
Seventy-six kindergarten children were recruited for this experiment. Parents received information on the study that was distributed by the kindergarten staff. Participation was voluntary and neither kindergarten nor parents nor children received any kind of monetary reward. Participating children received however little rewards after the experiment, for example small stickers.

Due to erroneous parental consent, data of one child was excluded from further analysis. Within this sample of 75 children, 11 reported growing up bilingual with German as their first language and a variety of languages as the second native language (French, Chinese, Greek), while three children reported another language (Dutch, Turkish and a Ghanaian language) as their native language. For none of these bilingual children any problems or difficulties in language comprehension were reported, hence data of all of these children was included in the analysis. The final sample, thus, consisted of 75 children (37 female) between the age of three and six years ($M_{age} = 4.29$ years, $SD = 0.834$).

Sample size was determined with reference to a previous study of our lab (Degner, in
DEVELOPMENTAL PERSPECTIVE ON LIB

preparation). Similar to the design of the here presented experiment, the respective study investigated how children acquired intergroup evaluations from storybook descriptions of two novel social groups in a similar design. Two conditions in this experiment introduced novel groups with nouns or adjectives labels and presented group members with a majority of positive vs. negative behaviors. Results demonstrated an interaction of group valence (“positive” group, i.e. the group depicted with mostly positive behavioral episodes vs. “negative” group, i.e. the group depicted with mostly negative behavioral episodes) and item valence (positivity vs negativity of the depicted behavioral episode). Reported effect sizes ranged between $\eta^2_p = .108$ and $\eta^2_p = .359$. Based on the smaller effect sizes, we aimed at collecting valid data from 30 children per condition for our experiment.

Unfortunately, the period of data collection overlapped with the term in which children typically transfer from kindergarten to elementary school, so that parts of the pre-recruited sample was no longer available at the time of data collection. This timing mishap was accompanied by additional difficulties in gaining parental consent in several neighborhoods. In combination these factors led to a loss of participants and a shortfall of the pre-determined sample size. Due to this premature termination of data collection, participants were not distributed perfectly even between conditions with $n = 25$ for the DAV, $n = 23$ for the IAV and $n = 27$ for the ADJ condition.

Yet demographic information of the participants did not differ significantly between groups, neither concerning age, $F(2, 74) = 0.262, p = .779$, nor concerning gender $\chi^2(2) = .107, p = .948$.

Dependent Variables.

To assess the potentially wide range of effects of language on the acquisition of group
representation, we employed a set of dependent variables, including the assessment of a) group evaluation, b) evaluative group traits, c) group preference, d) new member evaluation, e) new member preference or f) interpretations of ambiguous behaviors.

For the (a) group evaluation, we successively presented group pictures of the Lupis and the Nifos and asked “How nice…?” and “How naughty…?” they judged the entire group. Responses were given with the help of three-point smiley-scales, with smileys ranging from neutral to smile for the positive trait scale and ranging from neutral to a frown for the negative trait scale. An example is given in Figure 5. Additional to the visual of the smiley scale, the experimenter suggested three verbal response options while pointing to the corresponding icon on the scale. These verbal options were “not at all” (German: “gar nicht”, for the neutral smiley), “a little” (German: “ein bisschen” for the slight smile / frown) or “a lot” (German: “doll” for the big smile / frown.)

![Figure 5. Examples for the scale used to assess dependent variables.](image)

Children could give their answers by repeating the respective verbal response or pointing to the corresponding icon, or both.

For the measurement of (b) evaluative group traits, children were once more presented with pictures of both groups successively. Children were asked for each group to indicate, “How many Lupis or Nifos, respectively, are kind, nice, happy?” and “How many Lupis or Nifos, respectively, are naughty, mean, unfriendly?”.
Answering options were displayed visually by three different amounts of toy figures in the color of the respective group. Additionally, the experimenter verbalized response options and pointed to the respective amount of toy figures when uttering the respective verbal response option. These verbal response options were “none” (German: “gar keine”, “few” (German: “ein paar”) or “many” (German: “ganz viele”). Again, children gave their answers verbally, by pointing, or both.

For the measure of (c) children were shown pictures of both groups side by side, with the Lupis on the left and the Nifos on the right side, and asked to indicate with which group they would prefer to play with. Children answered verbally by giving the group label or pointing towards the respective group picture. In this experiment we used group preference as forced choice measure, i.e. we did not present children with the option to award both or none of them with a reward.

After completing the series of group evaluation variables, we assessed (d) new member evaluation. We therefore introduced participants to a single new member of each group. Here, we matched the new member’s gender to the participant’s gender. Similar to the assessment of the initial group evaluation, children were presented pictures of the new Lupi or the new Nifo, respectively, and asked to indicate “How nice…?” and “How naughty…?” they expected the new members of each group to be. Response options were identical to the ones used for measuring (a) group evaluation.

The (e) new member interaction preference was assessed analogously to the group interaction preference. That is, children were presented with pictures of one new member of each group side by side and asked to indicate with whom they would prefer to play with. Again, answers were given verbally or by gesture or both.
For exploratory reasons, we further included an adaptation of (f) an ambiguous behavior measure that was previously presented by McGlothlin, Killen and Edmons (2005). The ambiguous behavior test consisted of six additional pictures showing either a novel member of the Lupis or the Nifos in an ambiguously interpretable situation. Children were given a positive or negative interpretation of the depicted scene and were asked which scenario they perceived being more plausible. For instance, one picture showed a boy sitting in front of a large pile of toys and a girl sitting next to him. Participants were asked whether they believed that a) “the boy is intending to share all his toys” or b) “the boy possesses this number of toys, because he had taken them from other children”.

We implemented a counter-balanced design, such as each behavior was equally often shown by a Lupi or a Nifo, across participants.

The appendix contains a copy of the reporting sheet used by the experimenters including all dependent variables used in this experiment as well as an illustration of the ambiguous behavior task of Experiment 1.

Procedure.

Data collection was implemented from September 1st, 2015 until January 19th, 2016 and took place in five daycare-centers in the city center and suburbs of Hamburg, Germany.

Children were tested individually in the facilities of their daycare-centers with up to four children being simultaneously tested in one room. Prior to the beginning of the data collection, the experimenters introduced themselves to the participating kindergarten or during the so-called morning circle, for which all children attending the facility were gathered for the start of the day. After collecting the declarations of parental consent from the preschool teachers, participating children were led into a separate room in which several reading stations were prepared. Reading
stations were typically placed in the corners of the experimental room to reduce interaction between participants and held two child-sized chairs or cushions to allow participants to have a comfortable reading experience.

Each child was tested individually by one of five (four female, one male) experimenters, however not more than four experimenters were collecting data at the same time.

Assignment of conditions to experimenters was randomized. Our first step to ensure randomization was to let children choose the book they wanted to be read from to. We therefore designed all storybooks with an identical outward appearance and spread out the books in the middle of the room where participants picked their choice of storybook for the session. A second step to avoid demand effects, was to blind experimenters to condition. We therefore structured the test session in two phases: One experimenter read the story book to the children, another collected the dependent measures. Change of experimenter between phases seemed important to exclude an influence of demand effects on the side of either the experimenter or the children.

That is, we wanted to take precautions for the case that experimenters tended to steer the child’s answer in the hypothesized directions and for the case that children had the tendency to please the experimenter by simply repeating what he or she previously read from the storybook.

The first experimenter started the test session with an assessment of demographic variables and then made participants familiar with the storybooks. The experimenter initially introduced participants to the novel social groups and tried to ensure that participants correctly memorized the combination of group label and shirt color combination by repeating the mapping of color and label several times. Only afterwards, the experimenter read the story to the child. After the story was read to the participant, the participant switched to a second experimenters who then assessed the dependent variables in the above demonstrated fixed order from (a) group
evaluation to (f) ambiguous behavior task. Of course, this procedure implies that demographic information and dependent variables were noted on different response sheets and matched only by code number for data entry but at no phase during the experiment.

After finishing the second part of the experiment, i.e. the assessment of dependent variables, participants were thanked for their participation, chose a reward and were accompanied back to their kindergarten group.

Design.

Our experiment followed a 2 x 3 within-between design, consisting of the factors behavioral desirability (within) and linguistic category (between).

The first factor, behavioral desirability, arose from our manipulation of groups’ positivity or negativity, respectively. That is, throughout all conditions, the storybooks presented two novel groups, of which one was depicted with positive behaviors only, while the other was depicted with negative behaviors episodes only. Thereby, we created a perfect contingency between group membership and behavior desirability.

For the second factor, linguistic category, we manipulated the level of language abstraction of the narratives in the storybooks. Although the context descriptions were in DAVs for all narratives, we varied the abstraction with which the specific target behavior is described. In one condition all target behaviors were described in DAVs (e.g. “The Lupi gives every child exactly three marbles.”), in the second category all target behaviors were described in IAVs (e.g. “The Lupi is sharing the marbles with the other children.”), and in the third condition all behaviors are described in adjectives (e.g. “The Lupi is really generous.”)

Hypothesis.

In general, we expected children to acquire distinct evaluative representations of both
groups and these representations to be qualified by linguistic manipulation.

That is, firstly, we assumed that the group displayed with only positive behavior (nice group) would be evaluated more positively and less negatively compared to the group displayed with only negative behavior (naughty group) and that this effect would be mapped in interaction preferences, i.e. the nice group is preferred over the naughty group. We assumed that this effect would also hold for new group members. That is, we expected new members of the nice group to be evaluated more favorably compared to new members of the naughty group as well as interactions with new members of the nice group to be preferred over new members of the naughty group. However, evaluations of new group members could only be derived from evaluations of the nice group or the naughty group, respectively. For the ambiguous behavior task, we expected a main effect of group valence, i.e. we expected that ambiguous behavior of the nice group would more likely be interpreted as positive behavior and ambiguous behavior of the naughty group would more likely interpreted as negative behavior.

Secondly, we expected that differences in group evaluations and decisions for group preference would be qualified by linguistic manipulation, i.e. that abstract descriptions of behaviors – in contrast to concrete descriptions – would amplify the positivity or negativity in the behavioral descriptions, and thus generate more favorable or more unfavorable impressions of the groups. More precisely, we expected differences in group evaluations to be larger when children were exposed to more abstract narratives. Accordingly, we expected preferences for the nice group over the naughty group to be more profound when children were exposed to more abstract narratives. We built these expectations on research conducted with the Linguistic Category Model (Semin & Fiedler, 1988).

Although we assumed children of the full age-range to be susceptible to the linguistic
manipulation it is conceivable that children’s age moderates the expected effects, i.e. as older children have higher language proficiency their evaluations and preferences favoring the nice group over the naughty group might be even more pronounced compared to younger children.

**Results**

All statistical analyses were run with IBM\textsuperscript{©} SPSS\textsuperscript{©} Statistics (Version 23) for Mac. Results for each of the dependent variables as well as further exploratory analysis will be reported individually. Before implementing data collection, we were optimistic that all children within the age range were capable of closely listening to the storybook and would subsequently be able to conscientiously answer the questions measuring the dependent variables. However, practical experience proved us wrong. Children at the age of three years manifested a number of problems during the experimental procedure, e.g. attention deficits when listening to the story as well as during the assessment of the dependent variables, shyness regarding answering the experimenter and so on. Thus, we re-ran all analyses for this data while a) including and b) excluding three-year-olds. Since we found no differences when excluding compared to including the younger children, we subsequently report results analyzing the complete sample.

**Group evaluation.**

For analysis of group evaluation between conditions we conducted a 2 (group valence: nice group vs. naughty group) x 2 (item valence: positive evaluation vs. negative evaluation) x 3 (linguistic manipulation: DAV vs. IAV vs. adjective) repeated measures ANOVA. The analysis showed a significant interaction between group valence and item valence, $F(1, 71) = 8.067$, $p = .006$, $\eta^2_P = .102$, indicating that the nice group was consistently evaluated more positively and less negatively compared to the naughty group. Contrasting group evaluations in the three
linguistic conditions, we yet found no significant effect of the linguistic manipulation, $F(2, 71) = .357, p = .701, \eta^2_p = .010$. That is, group evaluation did not seem to depend on language abstraction. Figure 6 presents group evaluations in the three conditions.

Though we did not find the hypothesized effect of linguistic manipulation, we found children’s evaluations of the groups was indeed more positively for the nice group ($M_{Pos} = 2.66, SD = 0.603$) than for the naughty group ($M_{Pos} = 2.34, SD = 0.751$). A consistent pattern was found for the evaluation of group negativity with the nice group being evaluated less negatively ($M_{Neg} = 2.11, SD = 0.832$) compared to the naughty group ($M_{Neg} = 2.3, SD = 0.806$). An analysis of item valence and group valence illustrates that only for both of the rather concrete linguistic conditions (DAV: $F(1,23) = 9.104, p = .006, \eta^2_p = .284$ and IAV: $F(1,22), p = .031, \eta^2_p = .195$) the difference in evaluation is significant, whereas it is nonsignificant in the most abstract used condition (adjectives: $F(1,26) = .737, p = .398, \eta^2_p = .028$).

For further comparing group evaluations between conditions, we composed a difference score by subtracting the mean negative evaluation from the mean positive evaluation for both groups separately in each condition. Results for the nice group demonstrated that the combined evaluation score is positive for all conditions, i.e. DAV ($M_{PosEval-NegEval} = 0.45$), IAV ($M_{PosEval-NegEval} = 0.44$) and in the adjective condition ($M_{PosEval-NegEval} = 0.74$). The difference score for the naughty group attains a small negative value in the most concrete condition (DAV: $M_{PosEval-NegEval} = -0.08$) and only small positive values in the more abstract conditions (IAV: $M_{PosEval-NegEval} = 0.13$ and ADJ: $M_{PosEval-NegEval} = 0.08$).

While the numerical differences seem to indicate a more positive overall evaluation of the nice group compared to the naughty group as well as an increase in the evaluation positivity of the nice group for the more abstract conditions – which would be expected – and a decrease in
overall negativity for the naughty group in the most abstract conditions – which would be the contrary of what we expected - we cannot systematically interpret these results with regard to our hypothesis, since the effect of linguistic manipulation was not significant.
Figure 6. *Group based trait evaluation in the three experimental conditions of Experiment 1.*
Evalutive Group Traits.

A second approach to measuring group evaluation was the assessment of evalutive traits assigned to each of the novel social groups. For analyzing the evaluative traits, we first calculated an average evaluation of the positive (nice, kind, happy) and the negative (naughty, mean, unfriendly) trait scale, respectively.

A 2 (group valence: nice group vs. naughty group) x 2 (trait valence: positive trait evaluation vs. negative trait evaluation) x 3 (linguistic manipulation: DAV vs. IAV vs. Adjective) repeated measures ANOVA yielded a significant effect of trait valence, $F(1,71) = 34.413, p = 0.00, \eta^2_P = .0326$, indicating that both groups were overall evaluated more positively than negatively. Additionally, we found a significant interaction between group valence and trait valence, $F(1,71) = 11.734, p = 0.01, \eta^2_P = .142$, implying that more positive and less negative traits were awarded to the nice group than to the naughty group. However, this interaction was not further qualified by linguistic manipulation $F(2,71) = 0.921, p = .403, \eta^2_P = .025$, i.e. apportionment of evalutive traits did not depend on linguistic condition.

Comparing the evaluations of traits for both groups, replicated the pattern found for our initial measure of group evaluation. That is, the nice group was again evaluated more positively ($M_{Pos} = 2.25$) compared to the naughty group ($M_{Pos} = 2.04$) likewise the nice group was evaluated less negatively ($M_{Neg} = 1.63$) compared to the naughty group ($M_{Neg} = 1.84$).

Comparing results between conditions we find numerical differences contrary to our hypothesis, pointing in the direction that evaluations for the nice group are more positive in the most concrete (DAV: $M_{Pos Eval-Neg Eval} = 0.64$) and intermediate condition (IAV: $M_{Pos Eval-Neg Eval} = 0.67$) while lowest in the most abstract condition (ADJ: $M_{Pos Eval-Neg Eval} = 0.56$) and evaluations of the naughty group more positive with more abstract descriptions (DAV: $M_{Pos Eval-Neg Eval} = 0.07$, IAV: $M_{Pos Eval-Neg Eval} = 0.13$ and ADJ: $M_{Pos Eval-Neg Eval} = 0.36$). However since we found no effect of
linguistic manipulation, we cannot interpret this results with regard to our hypothesis.

Reliability for measurement scales was acceptable for consistent evaluations, i.e. evaluations on the negative trait scale for evaluations of the naughty group (Cronbach’s \( \alpha = .705 \)) and the evaluations on the positive trait scale for evaluations of the nice group (Cronbach’s \( \alpha = .614 \)). However, for seemingly inconsistent pairings of trait scale and evaluation, i.e. positive evaluations of the naughty group (Cronbach’s \( \alpha = .454 \)) and negative evaluations of the nice group (Cronbach’s \( \alpha = .310 \)) only poor levels of reliability were achieved.

Group Preference.

There was however no significant effect of linguistic condition on group preference, \( \chi^2(2) = 2.245, p = .326 \). 71 participants 47 (66.2 %) preferred an interaction with the nice group to an interaction with the naughty group. Four participants had refused to decide on an interaction preference for either group. Three of these four participants were assigned to the DAV and one participant to the ADJ condition. Distribution of interaction preferences between conditions is depicted in Figure 7.

![Figure 7](image_url)

Figure 7. Preference for groups in absolute numbers for the three experimental conditions of Experiment 1.
New Member Evaluation.

Analyzing group differences with a 2 (group valence: nice group vs. naughty group) x 2 (item valence: positive evaluation vs. negative evaluation) x 3 (linguistic manipulation: DAV vs. IAV vs. Adjective) indicated a significant effect of item valence, $F(1,70)= 19.269, p = .000, \eta^2_p = .216$, and a marginally significant interaction of group valence and item valence, $F(1,70) = 3.926, p = .051, \eta^2_p = .053$. However, there was no significant qualification by linguistic condition, $F(2,70) = 0.012, p = .998, \eta^2_p = .000$. Group evaluations were $M_{Pos} = 2.65$ ($SD = .67$) positive evaluations for the nice group and $M_{Pos} = 2.41$ ($SD = .81$) for the naughty group, and $M_{Neg} = 1.99$ ($SD = .87$) negative evaluations for the nice group, compared to the naughty group, $M_{Neg} = 2.12$ ($SD = .85$).

Analyzing new member evaluations for each linguistic conditions separately moreover yields no significant effects for the interaction of group and item valence neither for DAVs $F(1,23) = .213, p = .648, \eta^2_p = .008$ nor IAV: $F(1,22) = .107, p = .747, \eta^2_p = .005$ nor adjectives $F(1,23) = .107, p = 747, \eta^2_p = .005$).

While for the new naughty group members numeric differences hint at evaluations being more positive with more abstract descriptions (DAV: $M_{PosEval-NegEval} = 0.39$, IAV: $M_{PosEval-NegEval} = 0.08$, ADJ: $M_{PosEval-NegEval} = -0.14$), numbers do not represent a linear increase of positivity in evaluations for new nice group members (DAV: $M_{PosEval-NegEval} = 0.74$, IAV: $M_{PosEval-NegEval} = 0.43$, ADJ: $M_{PosEval-NegEval} = 0.77$) Moreover, since there is no significant effect of linguistic manipulation we cannot further compare respectively interpret numeric differences in group evaluation between conditions. Figure 8 contains an overview of new member evaluations in the three experimental conditions.
Figure 8. New member evaluation in the three experimental conditions of Experiment 1.
New Member Preference.

Interaction preferences did not differ significantly between conditions, $\chi^2(2) = .311, p = .856$. 73 participants overall 51 (69.86%) preferred an interaction with a new member of the nice group to an interaction with a new member of the naughty group. Two participants, both in the DAV condition, refused to decide on an interaction preference for new members. Distribution of interaction preferences between conditions is depicted in Figure 9.

![New Member Preference](image)

Figure 9. Preference for new members in absolute numbers for the three experimental conditions of Experiment 1.

Ambiguous Behavior Task.

To analyze how participants interpreted ambiguous behavior of groups, we first calculated how many ambiguous behaviors were interpreted in a positive respectively negative manner for both groups. A 2 (group valence: nice group vs. naughty group) x 3 (linguistic manipulation: DAV vs. IAV vs. Adjective) repeated measures ANOVA yielded a significant effect of group membership, $F(1,72) = 54.740, p = .000, \eta^2_p = .432$, with ambiguous behavior of the nice group being more likely interpreted in a positive way. However, again there was no significant effect of linguistic manipulation, $F(2,72) = 0.774, p = .456, \eta^2_p = .021$.

Results demonstrated that ambiguous behavior was more often subject to a positive
interpretation for members of the nice group ($M = 1.43, SD = .99$) compared to members of the naughty group ($M = 1.08, SD = 1.01$). Figure 10 illustrates the distribution of positive behavior interpretations between conditions.

![Ambiguous Behavior Task](image)

Figure 10. Averaged number of positive interpretations of ambiguous group behavior

**Discussion**

The aim of our first experiment was to test whether young children acquire evaluative representations of groups based on storybook narratives about group behavior and how linguistic abstraction in respective narratives influences the acquisition of group representations. For that purpose, we presented young children with a storybook introducing two novel groups. Both groups were presented with clear behavioral contingency. That is, presented were a nice group, whose members solely performed desirable behaviors, and a naughty group, whose members solely performed undesirable behaviors. Beside the valence of group behavior, we additionally varied the abstraction of storybook narratives from concrete only (DAV) over intermediate abstraction (IAV) to abstract (ADJ) descriptions.

We expected children to (1) acquire evaluative representations of both groups according
to the valence of presented group behavior and (2) these evaluative representations to be more pronounced when children were presented with more abstract storybook versions.

Consistent with our expectations on the first hypothesis, we found that children acquired evaluative representation of both groups in the intended direction. Namely, the nice group was rated more positively and less negatively compared to the naughty group on both measures of group evaluation. Correspondingly, children also expressed greater preference for an interaction with the nice group. Although our expectations were met for both evaluation measures, the effect sizes differed considerably: While we found a medium-to-large effect for group evaluation, the effect size for evaluative trait dimensions was small. This divergence of effect sizes might be due to the poor reliability of the measurement scale for the latter measure, which might have prevented accurate measurement of evaluative traits. Moreover, during data collection, there were visible signs of fatigue in children when asked for an assessment of the evaluative traits. Children seemed to be tired or confused by the repeated presentation of resembling questions. We also noticed that children had difficulties with the quantifying nature of the response scale, for example some children had difficulties to differentiate between “some” and “many”. It is therefore conceivable that children lacked language comprehension, cognitive skills or attention to process the distinct trait dimension and proper use them for group evaluation.

Results demonstrated that although children seemed to learn group evaluation with respect to group behavior and evaluated the group behaving only positively more positively and less negatively than the group behaving only negatively, there was no difference in group evaluations nor in any other dependent measure with regard to linguistic manipulations, i.e. the abstraction of language with which the behavior was described.

From these primary we derive several implications. First of all, we interpret these results
as a demonstration that the storybooks we used were suitable materials for our experimental purpose. Although the storybooks contained only a set of six narratives for each group, the limited amount of information we offered in the storybooks was already sufficient to prompt evaluation processes in children. Moreover, the materials we used were suitable for young children since children as young as four years were able to comprehend the presented narratives and to extract the social information within. We cannot, however, determine on which informational component children relied when generating group representations, i.e. whether children relied on visual information, behavioral descriptions or on both.

Secondly, these findings demonstrate how quickly children gather and process social information about novel groups. Based on our results, we argue that children are not passive recipients of social information; rather, they actively use social information to construe an understanding of social groups from a young age onward. Accordingly, our results can be an indication for the importance of communication for the formation of intergroup attitudes.

In a second step, we analyzed whether children transferred the information they received about groups and the thereupon established group evaluations to new group members. For that purpose, we measured evaluations of and preference for new group members. We found that new members of the nice group were being evaluated more positively and less negatively compared to new naughty group members. However, the evaluative difference between the new group members was only marginally significant and the effect size was small. This may indicate that children do not to a full extent transfer evaluations of groups to new and unknown members.

For further assessment of the potential effects on group evaluation, we presented children with storybook episodes presenting behavior that was ambiguously and thus in need of interpretation. Children were then asked to judge whether the actor performed a positive or
negative behavior. The results clearly showed that children interpreted ambiguous behavior of the nice group in a positive way more often than for the naughty group and that this effect was quite large. Considering the perfect contingency between group membership and behavioral valence, the magnitude of the effect is not surprising. A possible inference from this finding is, that children use social information not only to quickly build evaluative group representations but also to build expectations towards these groups. We can only presume at this point, that these expectations might also influence future interactions with respective groups.

The second purpose of our experiment was to investigate how language abstraction influences the acquisition of group representations. We assumed that representations of groups would be distinctly more positive respectively more negative when children were presented with abstract compared to concrete behavioral descriptions. Thus, we manipulated the abstraction of narratives and presented children with storybooks using either DAVs, or IAVs or adjectives, only. However, we found no significant effect for linguistic manipulation on any of the reported measures. That is, while children seemed to rely on the valence of behavior to form representations of groups, an increasing abstraction of the behavioral descriptions did not seem to assert additional influence on group evaluation. There are however several aspects of our experimental design, which are worth considering to be at least partially responsible for the lacking effect of linguistic manipulation.

Firstly, it seems possible that children might process word categories somewhat differently from adults and results therefore deviate from our hypothesis. From research on language development in childhood, we know that adjectives for language acquisition become important rather late. During elementary school years, with a total of 11% adjectives only represent a small proportion of children’s vocabulary (Rössl, 2007). It is conceivable that the
unfamiliarity of our participants with adjectives as a linguistic category as such, inhibited the unfolding of a potential abstraction effect.

Secondly, the outline of our storybook and the presentation of the novel groups were straightforward. Throughout the narratives we created a perfect contingency between group membership and behavior type, forthrightly introducing one group as the nice one, the other group as the naughty one. This simplicity of the narratives possibly helped children to categorize and built distinct evaluations of the novel group, but maybe interfered with the experimental purpose on another level. It is conceivable, that the association we created between group membership and behavior type was so strong, that the influence of subtle linguistic manipulations was leveled out. In other words, processing of social information might be subject to influences of implicit linguistic connotations in communication only when the information is to some extent ambiguous or uncertain. Strong association between group membership and behavioral valence is also problematic when considering ecologic validity. In real life social settings, it is rather unlikely that children receive information about social groups bare of any ambiguity.

Summarizing the results of our first experiment, we can conclude that children quickly acquire evaluative representations of novel groups based on behavioral information. Children further express according interaction preferences and, to a limited extent, transfer the acquired group evaluations to new members of the respective groups. The abstraction of language used for behavioral description, however, does seemingly not influence the acquisition of group representations when behavioral information about groups unequivocally indicates positivity or negativity.
Experiment 2: Effects of Linguistic Differences in Behavioral Descriptions on the Acquisition of Intergroup Evaluations by Preschool Children\(^2\)

Based on the results of our first experiment, we altered several aspects of the study’s design to further analyze how linguistic intergroup biases affect the formation of novel group attitudes in young children.

Firstly, we did not include three-year-olds in our sample, because we previously experienced that language comprehension and attention span in these children were typically not sufficient to follow the complete experimental procedure. Moreover, instead of solely focusing on pre-school children, we extended our sample adding elementary school children. The inclusion of older children should give us a possibility to test whether the absence of linguistic effects could be attributed to lack of language comprehension and processing skills of younger children.

Secondly, we took a slightly different angle on our main research question and adjusted essential features of the experimental manipulation. Although, we again presented storybooks introducing two novel social groups, we dismissed behavioral consistencies for both groups. Thus, instead of presenting the groups again only performing solely positive or negative behaviors, we now presented both groups behaving equally often in desirable and undesirable ways. That is, we balanced the ratio of positive and negative behavior in order to create an ambiguous behavioral representation of both groups. Accordingly, as both groups were displayed with the same amount of positive and negative behavior, there is no obvious difference between those groups. Consequently, based on the content of behavioral information about groups,

\(^2\) We preregistered our experiment with the Open Science Framework, the registration is retrievable under https://osf.io/c9rt7/
evaluations of both novel groups should be equal or at least similar.

In order to create the perception of a rather nice respectively a rather naughty group, despite the absence of an actual difference in demeanor, we used variation of language abstraction in behavioral descriptions. For one group, desirable behaviors are described using abstract terms and undesirable behaviors are described using concrete terms (Group1), for the other group, desirable behaviors are described using concrete terms and undesirable behaviors are described using abstract terms (Group2).

Thirdly, we extended our linguistic manipulation to include nouns as the fourth and most abstract condition. Even though nouns are not part of the LCM originally, subsequent research on potential extensions of the LCM has yielded results favoring an integration of nouns as most abstract category into the taxonomy of the LCM; Carnaghi et al. (2008).

Nouns are typically used to categorize, i.e. referring to someone with a noun places him or her in a certain class of people. By this act of categorization all information and knowledge about the respective category, is typically ascribed to the categorized subject. Moreover, nouns are believed to convey a complex set of information that goes beyond the pure sum of specific traits. Imagine your mental representation of someone described as very smart, rather introvert, socially shy and not very athletic compared to the image evoke by the label “IT-Nerd”. Whereas the former adjectives are all part of the “nerdy stereotype”, using the label itself probably elicits a representation well beyond the actual listed qualities, perhaps of someone wearing glasses, likely male, someone who tends to spend his Friday night alone gaming while eating pizza right out of the box. This is an example on how the use of nouns for interpersonal descriptions does not only contain more information but produces a lot more associations compared to adjectives which only assign a specific quality to the described person.
Adjectives and nouns moreover differ regarding their exclusivity. Whereas qualities assigned by adjectives can be relative, e.g. someone can be very aggressive and somewhat musically, nouns feature an either-or-quality, i.e. either someone is a hooligan or a musician or none of the two, respectively.

Based on these linguistically determined differences between adjectives and nouns, Carnaghi et al. (2008) hypothesized that the use of a noun for description would associate the subject of description with plenty of qualities attached to this noun. Consequently, the use of nouns would produce stronger, stereotype-congruent inferences. Carnaghi et al. (2008) postulate moreover, that the use of nouns inhibits alternative classifications of the described subject and thus conveys greater essentialism, i.e. the use of nouns is considered to fuel beliefs that root the existence of qualities to the nature of a person rather than to external circumstances.

Comparing this information on the effects of nouns with the known effects of adjectives on impression formation, demonstrates that nouns to some extent seemingly assert stronger influences on impression formation. More precisely, the effect of nouns exceeds the effect of adjectives on the dimension of perceived subject-related informativeness. Regarding the enduringness of qualities ascribed by nouns, they moreover seem to have a more tangible base for prediction on future events.

While these results indicate that nouns could indeed fit into the logic of the LCM and extend it to a fifth category, the research of Carnaghi et al. (2008) as well produced contradictory evidence. Nouns were, for example, found to easily elicit vivid images of the context in which they are used, a quality that is typical for concrete rather than abstract linguistic categories. It is assumed that the use of nouns activates prototypical exemplars of the respective category, like for example the noun “nerd” might by some be associated with the character of Sheldon Cooper
on the popular television show “The Big Bang Theory”.

Thus, even though nouns do not surpass the effect of adjectives in all dimensions regarding person perception, Carnaghi et al. (2008) provided empirical evidence regarding the effect of nouns on impression formation that justifies to treat nouns as a fifth, and the most abstract category, of the LCM. We therefore assume that – given there is a causal influence of language abstraction on impression formation – nouns might assert the strongest influence and consequently included nouns into the linguistic manipulation of our second experiment. That is, we used a linguistic manipulation as follows: While, for concrete behavioral descriptions, we relied entirely on DAVs, we varied the level of abstraction for the abstract behavioral conditions. That is, we used IAVs in one condition, adjectives in the second condition, and nouns in the third condition, thus, increasing the degree of abstractness from the first to the third condition (see Coenen, Hedebouw and Semin, 2006).

Based on the available literature on linguistic intergroup biases (Carnaghi et. Al., 2008; Maass et al., 1989; Salès-Wuillemin et al., 2014; Schoel et al., 2014; Semin & Fiedler, 1988), we expected the more abstractly described behaviors to have a stronger influence on novel group representation. That is, if a group was presented with abstract positive behaviors and concrete negative behaviors, we expected the group to be evaluated more positively and less negatively. In contrast, if a group was presented with abstract negative behaviors and concrete positive behaviors, we expected the group to be evaluated more negatively and less positively.

We expected this effect to occur based on our linguistic manipulation of behavioral descriptions only since group’s behavior was lacking clear positive or negative valence. We further expected representations of groups to be more pronounced the larger the contrast of abstraction in the narratives. Thus, we expected differences in group evaluations to be smallest
when positive behavior of Group 1 is described in IAVs, while positive behavior of Group 2 is described in DAVs, larger when positive behavior of Group 1 is described in adjectives, while positive behavior of Group 2 is still described in DAVs and largest when positive behavior of Group 1 is referred to with nouns, while positive behavior of Group 2 still described in DAVs.

**Method**

**Materials.**

Storybooks were overall similar to Experiment 1. On the first page of the storybook, children were introduced to both groups, i.e. the Lupis and Nifos, and children were made aware of group labels and the t-shirt color indicating either group membership. These group features were repeated up to four times to ensure children’s proper and distinct identification of both groups.

The storybook consisted of twelve behavioral episodes, in which we presented only intragroup behaviors. However, different from our previous experiment, we balanced positivity and negativity for the behavior of both groups. That is, we evenly presented both groups with three socially desirable and three socially undesirable behaviors. Thus, there was no consistency of the group members’ overall behaviors. For the behavioral episodes we re-used five scenes from our first storybook (protecting, excluding, sharing, disobedience, consoling), yet we had to slightly alter the facial features of a group member in these episodes, because children perceived the form of his eyebrows as an indication for aggression. We further used three episodes displaying contrary actions to those used in the first storybooks (tidying up, being irresponsibly loud, cleaning oneself up) and added four new behaviors (aggression, cheering up, not sharing/being greedy, recklessness). The added behaviors were as well chosen to be common and relevant for the participants’ age group.
We aligned the formal features with our previous storybook design. That is, all episodes were presented equally on a double-page, containing two pictures sized 13 x 16 cm, with the situational context established on the left and the concrete behavior displayed on the right page. Whereas narratives of the situational context were identically worded DAVs across all conditions, descriptions of the focal actions varied in the manner that they either contained an IAV, a person-describing adjective, or a person-describing noun. An example for a storybook page is given in Figure 11. For the example behavior would be described with phrases that translate to either “The Nifo hurts the other child.” (IAV), or “The Nifo is mean.” (Adjective) or “The Nifo is a villain.” (Noun).

![Figure 11. Example episode from the storybook in Experiment 2.](image)

Participants.

We altogether recruited 207 participants; from those, 97 attended kindergarten and 110 attended the first two grades of primary school, respectively. Notwithstanding our pre-defined age group criteria, in some facilities kindergarten teachers recruited children aged younger (n=4)
than three years. To prevent these children from experiencing disappointment and exclusion, we conducted the experiment regardless and excluded their data from analyses. Within our elementary school sample, the experimenters reported insufficient German language skills for three children, whose data was subsequently excluded from analyses as well.

Altogether, 27 children self-identified as bilingual, from which three reported a native language different from German, namely Arabian, Russian and Serbian. As there were no comprehension problems reported for these bilinguals, these data sets remained in the sample.

Thus, our final sample consisted of 200 children (46.5% kindergarten children). The age ranged from 4 to 11 years with $M_{age} = 6.52$ years ($SD = 1.93$). 53% of participants were male.

In the kindergarten subsample 50.5% of participants were male, children were 4.81 years old on average ($SD = .74$), respectively in the elementary children subsample 55.1% of participants were male, with $M_{age} = 8.00$ years ($SD = 1.3$).

For the purpose of determining the sample size we again relied on the previously cited study conducted in our lab (Degner, in prep.). Accordingly, we aimed at including at least 30 participants per condition. Our final sample was distributed comparably between conditions, with $n = 38$ for the first, $n = 37$ for the second and $n = 35$ for the third condition. Again, demographic variables of the participants did not differ significantly between groups, neither concerning age, $F(2, 197) = .987, p = .374$ nor concerning sex $\chi^2(4) = 4.388 p = .356$.

Dependent Variables.

Children’s perceptions of both novel groups were measured in three steps containing a) group evaluation, b) reward allocation, c) group preference, d) new-member evaluation, e) new-member-preference and f) an ambiguous behavior.

For (a) the measure of group evaluation, we extended the variables set used in
Experiment 1 by additionally referencing groups to exemplary positive and negative behaviors and general liking. That is, for the group evaluation measure we asked children (1) “How often do the Lupis/Nifos help other children?”, respectively “How often do Lupis/Nifos irk other children?”, (2) “How nice are the Lupis/Nifos?” respectively “How naughty are the Lupis/Nifos?” and (3) “How much do you like the Lupis/Nifos?”.

Group evaluation was, thus, assessed on several dimensions of abstraction by using IAVs for a more behavioral orientated evaluation (1), adjectives for a trait-based evaluation (2) and state verbs to determine general liking of both groups (3). We assumed, that extending the evaluative dimensions of this measure would allow a more comprehensive evaluation of groups and presumed that initiating this measure with a concrete behavioral example (1) would make this measure more graspable and accessible to children. The evaluative questions were once more presented as a question block for each group.

Likewise, as we did in Experiment 1, we gathered responses for all questions on three-point-smiley-scales, each point marked by a smiley face and an accompanying quantifier. Smiley faces ranged from neutral to a big smile for the positive dimensions and the general liking scale and from neutral to frowny face for the negative dimensions. Accompanying verbal quantifiers were “never” (German: “nie”), “sometimes” (German: “manchmal”) and “often” (German: “oft”) for the evaluation of behavioral tendencies, e.g. helping and irking, respectively “not at all” (German: “gar nicht”), “a little” (German: “ein bisschen”) or “a lot” (German: “doll”) for the trait dimension scale, e.g. being nice and naughty as well as for the assessment of general liking. Children indicated their response either verbally or by gesturing to the appropriate smiley.

Instead of measurement of evaluative group traits used Experiment 1 we employed a (b) reward allocation task in Experiment 2 as a second measure of group evaluation. Reward
allocation strategies have previously reflect linguistic biases (Maass et al., 1996). We expected this measure to be more child-friendly concerning task difficulty and attention demands in comparison to the evaluative trait measure of Experiment 1.

We chose four different child-appropriate rewards, i.e. cookies, presents, toys and ice-cream. On four consecutive pages, we presented a group picture of either the Lupis or the Nifos and three pictures of the respective reward, showing a small, medium and large version of the reward. Children accordingly had to choose between a small, medium or large number of cookies; a small, medium and large sized version of an identically wrapped present; between differently valuable toys, i.e. a crayon (small reward), a sand bucket with shovel (medium) and a pedal car (large) and finally between an ice-cream cone with one (small), two (medium) or three (large) scoops of ice-cream. On each page, children were given a short explanation in which context the reward should be distributed. None of these introductory remarks prompted children to directly associate previous group behavior and reward distribution. Reward allocation questions were presented in a set of four for each group individually. Children indicated their responses, again, either verbally or by pointing towards the respective reward.

Anew, we assessed (c) group preferences by presenting pictures of both groups on a double-page, with the Lupis on the left and the Nifos on the right side. In contrast to Experiment 1, children could, however, not only indicate their preference to interact with the Lupis or the Nifos, but as well indicate a preference for an interaction with both or none of the groups.

We complemented these series of directly group-related dependent variables with the evaluation of (d) new members of each group. New member evaluation was again aligned to the group evaluation measure, thus granting us the opportunity to test whether children transferred group-based evaluations to novel and unknown group members. We therefore presented children
with the picture of a new group member of the same gender as the participant and subsequently asked “How often does this new Lupi/Nifo help other children?”, respectively “How often does this new Lupi/Nifo irk other children?”; b) “How nice is this new Lupi/Nifo?” respectively “How naughty is this new Lupi/Nifo?” and c) “How much do you like this new Lupi/ Nifo?”. Response options were identical to the first group evaluation measure, i.e. presented with three-point smileys scales, supplemented by verbal qualifiers as specified above. Questions were, again, presented in separate sets for members of each group.

Children were asked for (e) new member preference while being presented to pictures of both groups. Same as for group preferences, response options were supplemented by the choices “preference for both” and “no preference”, respectively.

We concluded our set of dependent variables with an (f) ambiguous behavior task. For this task, we presented children with five ambiguous situations, in which either a Lupi or Nifo acted in a positive or negative way. Depicted situations displayed neither intra- nor intergroup contexts. That is, the target of an action was referred to as “this child” instead of being labeled with a group name. Additionally, the clothes of the “target child” were colored in grey and did not relate to the respective group colors blue or orange.

The ambiguous situations were presented in a single picture for which we provided a short situational context. The experimenter then gave a standardized positive and negative interpretation of the picture and asked children, which of the given interpretations was more likely to explain what happened or is going to happen in the picture. For instance, in one picture a Lupi or Nifo is sitting in a sand box and has a pile of toys in front of him, while a child sitting next to him has no toys. Participants were made aware of this situation and subsequently asked whether the target Lupi or Nifo is (a) about to share his toys with the child who has nothing or
(b) has taken the toys from the child. Within the sequence of pictures, we varied whether the positive or the negative interpretation was offered first.

We presented the set of five ambiguous behaviors once for each group, i.e. the identical five ambiguous situations with a Lupi as the actor and, subsequently, with a Nifo as the actor.

Procedure.

Data collection for the kindergarten sample was realized between March 15th, 2016 and May 31st, 2016. The experiments were conducted in five kindergartens in the city of Hamburg. Due to local limitations of a required governmental permit, data collection for the elementary school sample did not take place in Hamburg, but in the federal state of Schleswig-Holstein. Data for this school sample was collected from December 6th, 2016 to December 9th, 2016 in a single school in Schleswig-Holstein.

The experimental procedure was identical for kindergarten and elementary school children. The leading experimenter initially collected parental consent from the responsible teacher and met participating children in their kindergarten group or classroom. The experimenters introduced themselves, led children to a separate room, in which the experiment was conducted. During the entire process overall six female and one male experimenter were involved in data collection, although not more than four experimenters simultaneously partook in one test session. Upon entering the room, children were invited to choose one of the picture books spread out in the middle of the room. Letting children pick from a stash of identically looking books was again used to randomize distribution of condition between participants. However, randomization was limited by our need to equalize distribution of participants between conditions. We therefore regularly counted how many participants were assigned to each condition. In case we found a larger imbalance, we omitted the picture books that were already
read above average from the stash children could choose from for the next couple of test sessions.

Apart from that, experimental procedure was held closely to the proceedings in Experiment 1. After choosing a book, children were invited to sit next to one of the experimenters. The fist experimenter assessed a small set of demographic variables, i.e. age, gender, native language, as well as grade if applicable, and noted this information together with the experimental condition and the child’s participation number on a documentation sheet. The experimenter then introduced the Lupis and Nifos and tread the storybook to the child. After listening to the story, children were encouraged to switch seats and a second experimenter, blind for the experimental condition, assessed the dependent variable. Results were again noted on a separate reporting sheet and by matching subject numbers only later complemented with demographic information and condition assignment.

Dependent variables were collected in the order as presented above: a) group evaluation, b) group preference, c) reward allocation, d) new member evaluation, e) new member preference and f) ambiguous behavior task. Children were afterwards thanked for participation and rewarded with a small present, i.e. kindergarten children chose from a variety of stickers and elementary school children from a large set of colored rubbers.

Design.

Experiment 2 likewise follows a 2 x 3 within-between design. For the within-factor we linguistically manipulated perception of groups. Since we presented both groups with an equal amount of positive and negative behavior, there was no objective distinction of group valence. We, instead, relied on a specific linguistic pattern for behavioral descriptions to create the impression of group positivity, respectively group negativity. That is, we described positive
behaviors of Group 1 abstractly and negative behavior concretely. For Group 2 we reversed the
description pattern, that is, we described negative behaviors of Group 2 abstractly and positive
behavior concretely.

For the second factor, linguistic category, we manipulated on what level of abstraction
behaviors were described. We, however, varied only how abstract positive behavior of the nice
group and negative behavior of the naughty group are described. Accordingly, in the first
condition, we described positive behavior of Group 1 and negative behavior of Group 2 with
IAVs, in the second condition we used adjectives, and in the third condition we used nouns. The
latter have been previously rated by a number of kindergarten teachers to ensure children would
be familiar with the used nouns.

The remaining behaviors, i.e. negative behavior of Group 1 and positive behavior of
Group 2, was given in DAVs as well as context descriptions.

Hypothesis.

We basically expected children to acquire distinct evaluative representations of both
groups based on linguistic descriptions of their behavior, although both groups were presented
equally positive and negative. That is, we expected children to acquire different evaluations of
groups based only on the linguistic pattern of behavioral descriptions. Building on research of
the LCM (Semin & Fielder, 1988) and the LIB (Maass et al., 1989), we assumed that abstractly
described behaviors would be perceived as more stable and more diagnostic. Consequently,
although there is no difference between groups in behavioral positivity and negativity, the
description pattern we used, could generate a more favorable impression of Group 1, for which
positive behavior was described abstractly, and a rather unfavorable impression of Group 2, for
which negative behavior was described abstractly.
Rephrased in the terms of Experiment 1: We expected that Group 1 is perceived as the rather nice group and Group 2 is perceived as the rather naughty group. We expected according results for group preference as well as reward allocation. In our first experiment we could demonstrate that children to some extent transfer group evaluations to new member evaluations. We assumed that this tendency would be apparent in our second experiment as well, thus we concluded that new members of Group 1 would be evaluated more positively compared to members of Group 2 – yet we acknowledged the possibility that new member evaluations might be be less pronounced compared to group evaluations. Finally, we expected our manipulation to affect results of the ambiguous behavior task in a similar way, i.e. we presumed ambiguous behavior of Group 1 to be more likely interpreted in a positive way.

These hypotheses all referred to the manipulation of the first experimental factor, i.e. the type of behavior described in an abstract manner for each group. Our second manipulation was also linguistically based but referred to the grade of abstraction in behavior descriptions. Based on the effects of language affection on impression formation (Semin & Fiedler, 1988), we expected group perception to be more pronounced when more abstract categories were used for descriptions. Concretely, that means, that Group 1 would presumably be perceived more positively in the more abstract the linguistic category used for description, while Group 2 would presumably be perceived more negatively the more abstract the categories used for behavioral descriptions. Accordingly, evaluative difference between groups is expected to be larger in the more abstract conditions. We expected this abstraction effect for all dependent measures, i.e. group evaluation, reward allocation, group preference, new member evaluation, new member preference and interpretation of ambiguous behavior.
Results

All statistical analyses were conducted with IBM® SPSS® Statistics (Version 23) for Mac. Results for each of the dependent variables, as well as the exploratory analysis, will be reported individually. Based on our experience with kindergarten children in Experiment 1, we wanted to secure that deficient language comprehension and attention span did not interfere with the experimental manipulation, thus included elementary school children in our sample. We therefore considered sample, i.e. kindergarten vs. elementary school, an additional factor in our analysis. If results for our age-related subsamples significantly diverged, we report results for sample groups separately.

Group evaluation.

We measured group evaluation on different levels of abstraction, with three questions assessing a positive evaluative dimension (helping, being nice, general liking) two questions (irking, being mean) assessing the negative evaluative dimension.

To test our hypothesis that abstract descriptions amplify differences in group evaluation, we analyzed the influence of abstraction by including linguistic manipulation as a factor in a 2 (group valence: nice group vs. naughty group) x 2 (trait valence: positive evaluation vs. negative evaluation) x 3 (linguistic manipulation: IAV vs. Adjective vs. Noun) repeated measures ANOVA. This three way ANOVA demonstrated no qualifying effect of linguistic manipulation, $F(2, 196) = 0.171, p = .843, \eta^2 = .002$. However, we found an interaction between group valence and linguistic manipulation with $F(2,196) = 3.478, p = .033, \eta^2 = 0.034$. The difference in group evaluations was significant in the most abstract condition only, $F(1, 64) = 7.429, p = .008, \eta^2 = .104$. That was true for both sample groups, that is, group evaluation differed significantly only in the adjective conditions in the preschool sample with $F(1,28) = 4.492, p =$
.043, η²_P = .138 as well as in the elementary school sample with F(1,28) = 4.492, p = .043, η²_P = .138 yet no significant differences were found when using IAVs neither in the preschool sample with F(1,32) = 0.743, p = .395, η²_P = .023 nor in the elementary school sample with F(1,33) = 0.440, p = .512, η²_P = .013 or when using adjectives with F(1,29) = 0.021, p = .887, η²_P = .001 for the preschool sample and F(1, 36) = 0.00, p = .953, η²_P = .000.

We further analyzed the effect of sample to detect potential age effects and found a significant interaction between item valence and sample, F(1,193) = 10.643, p = .001, η²_P = .052, as there was a main effect of item valence for elementary school, F(1,89) = .006, p = .983, η²_P = .000 but not for pre-school children, F(1,104) = 32.761, p = .000, η²_P = .240.

For our analysis, we averaged scores on each dimension for both groups, and used these aggregated values as indices for positive, respectively negative evaluations. Yet, it is important to note that reliability for the evaluative dimension scales was altogether poor, with α = .532 (Group 1, the “abstract positive group”) respectively α = .521 (Group 2, the “abstract negative group”) for the positive evaluation scale and α = .543 (Group 1, the “abstract positive”) respectively α = .500 (Group 2, the “abstract negative”) for the negative evaluation scale. For an overview see Table 3.

Table 3. Reliability of the positive and negative evaluation scales for group evaluation in Experiment 1.

<table>
<thead>
<tr>
<th></th>
<th>Group 1 (“abstract positive”)</th>
<th>Group 2 (“abstract negative”)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Positive Evaluation</td>
<td>Negative Evaluation</td>
</tr>
<tr>
<td>Pre-School</td>
<td>α = .604</td>
<td>α = .425</td>
</tr>
<tr>
<td>Elementary School</td>
<td>α = .401</td>
<td>α = .656</td>
</tr>
</tbody>
</table>
Figure 12 illustrates differences in group evaluation between conditions in the sample of pre-school children; Figure 13 illustrates respective differences in the sample of elementary school children.
Figure 12. Group evaluation in the preschool sample for the three experimental conditions of Experiment 2.
Figure 13. Group evaluation in the elementary school sample for the three experimental conditions of Experiment 2.
Group preference.

We found no difference for group preference between conditions with $\chi^2(6) = 1.1660, p = .948$ neither differences between sample groups $\chi^2(3) = 2.178, p = .437$. Overall 98 out of 200 participants preferred to interact with the nice group (49 %), 12 participants indicated a preference for both and 5 participants for none of the groups. Distribution of interaction preferences between conditions is depicted in Figure 14.

![Group Preference](image)

**Figure 14.** Preference for groups in absolute numbers for the three experimental conditions of Experiment 2

Reward Allocation.

To analyze the distribution of rewards we summarized responses on the four reward items and calculated an average of reward allocation for both, the nice and the naughty group.

A 2 (reward allocation: Group 1: “abstract positive” vs. Group 2: “abstract negative”) x 3 (linguistic manipulation: IAV vs. Adjective vs. Noun) repeated measures ANOVA yielded a marginally significant effect of group valence, $F(1,199) = 2.802, p = .094, \eta^2 = .014$, indicating a greater distribution of rewards to Group 1 (“abstract positive”). Yet, this effect was not further qualified by linguistic manipulation, $F(2,197) = 0.318, p = .728, \eta^2 = .003$.

Reward allocation did not differ significantly for neither IAV use with $F(1,67) = 0.440, p$
Developmental perspective on LIB

= .509, η² = .007 nor use of adjectives with $F(1,66) = .502, p = .481, \eta^2 = .008$ nor use of nouns with $F(1,64) = 2.122, p = .150, \eta^2 = .032$. However, it seems notable, that only for the most abstract condition (nouns) differences in reward allocation between groups at least approximates the level of significance. Figure 15 shows the allocation of rewards between conditions.

Analyzing for sample effects with a 2 (reward allocation: nice group vs. naughty group) x 2 (sample group: kindergarten vs. elementary school) x 3 (linguistic manipulation: IAV vs. Adjective vs. Noun) ANOVA yielded no significant effects, $F(2,194) = 0.883, p = .415, \eta^2 = .009$.

Reliability for the reward item scales was acceptable, with $\alpha = .734$ for the Group 1 (abstract positive) and $\alpha = .751$ for Group 2 (abstract negative). Consistent with our hypothesis, children awarded more rewards to Group 1 (“abstract positive”) $M = 2.25$ ($SD = 0.52$) compared to Group 2 (“abstract negative”) $M = 2.17$ ($SD = .55$).

Figure 15. Distribution of rewards in the three experimental conditions of Experiment 2.
New member evaluation.

To assess new member evaluation, we employed the same scales used for initial group evaluation. In like manner, we averaged items on the positive respectively negative scale and used the results as aggregated evaluation scores. Mean scores for group evaluations are demonstrated in Table 4.

Conducting a 2 (group valence: nice group vs. naughty group) x 2 (trait valence: positive evaluation vs. negative evaluation) x 3 (linguistic manipulation: IAV vs. Adjectives vs. Noun) repeated measures ANOVA we did not found an effect of linguistic manipulation, $F(2,195) = 0.295, p = .745$.

<table>
<thead>
<tr>
<th>Table 4. Mean positive and negative evaluations of new group members in Experiment 2.</th>
</tr>
</thead>
<tbody>
<tr>
<td>New member of Group 1 (”abstract positive”)</td>
</tr>
<tr>
<td>$M$</td>
</tr>
<tr>
<td>Positive Evaluation</td>
</tr>
<tr>
<td>Negative Evaluation</td>
</tr>
</tbody>
</table>

To further investigate potential sample effects, we conducted a 2 x (group valence: nice group vs. naughty group) x 2 (trait valence: positive evaluation vs. negative evaluation) x 2 (sample: pre-school vs. elementary school) repeated measures ANOVA. This ANOVA yielded a significant interaction between group valence and sample, $F(1,196) = 4.920, p = 0.028, \eta^2 = 0.24$.

Analyzing the sample effect in conditions one by one, we found a significant interaction between group valence and sample group only in the adjective condition $F(1,64) = 4.055, p = 0.048, \eta^2 = .060$. In this condition, pre-school children evaluated the “abstract negative” more
positively and more negatively than the nice group, while elementary school children evaluated the nice group marginally more positively and less negatively compared to the naughty group.

Neither for the linguistic manipulation with IAVs with $F(1,65) = 1.661, p = .202, \eta^2_p = .025$ nor with nouns with $F(1,65) = 0.229, p = .634, \eta^2_p = .004$ did the interaction of group valence and sample valence yield significance.

We moreover found a significant interaction of item valence and sample in the noun condition $F(1,63) = 6.272, p = .015, \eta^2 = .091$, respectively marginally significant in the adjective condition, $F(1,64) = 2.888, p = .094, \eta^2 = .041$ whereas this interaction was not different in the IAV condition $F(1,65) = 7.21, p = .399, \eta^2_p = .011$ These results reflect that both groups are evaluated distinctly more positive than negative. Figure 16 presents group evaluation given by both sample groups in the three experimental conditions.

Reliability of scales approximated acceptable levels of reliability (see Table 5).

*Table 5. Reliability of the positive and negative evaluation scales for new member evaluation in Experiment 2.*

<table>
<thead>
<tr>
<th>Group 1 (“abstract positive”)</th>
<th>Group 2 (“abstract negative”)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Positive Evaluation</td>
<td>Negative Evaluation</td>
</tr>
<tr>
<td>Pre-School</td>
<td>$\alpha = .713$</td>
</tr>
<tr>
<td>Elementary School</td>
<td>$\alpha = .694$</td>
</tr>
</tbody>
</table>
Figure 16. New Member evaluation for both groups in the preschool and elementary school sample of Experiment 2.
New member preference.

Preference for new members was not influenced by linguistic manipulation $\chi^2(6) = 4.356, p = .629$ neither did effects of sample occur $\chi^2(3) = 3.710, p = .295$. 90 out of 195 participants signaled a preference for the “abstract positive” over the “abstract negative” group, while 18 participants indicated a preference for an interaction with both groups and five participants preferred to have no interaction with either of the groups. Distribution of interaction preferences between conditions is depicted in Figure 17.

![Figure 17. Preference for new members in absolute numbers for the three experimental conditions of Experiment 2](image)

Ambiguous behavior task.

For analysis of the ambiguous behavior task we summarized how many ambiguous behaviors were interpreted in a positive way for each group.

A 2 (group valence: nice group vs. naughty group) x 3 (linguistic manipulation: IAV vs. Adjective vs. Noun) repeated measures ANOVA yielded no significant effects, i.e. children did not interpret ambiguous behaviors for groups differently, $F(1,195) = 0.007, p = 0.933, \eta^2 = .000$, nor did interpretation of group behavior depend on condition, $F(2,195) = 0.747, p = 0.475, \eta^2 = .008$. On average children interpreted $M = 2.90$ interactions of the “abstract nice” ($SD = 1.54$)
and of the “abstract negative” group as positive ($SD = 1.50$).

Analyzing interpretation of ambiguous behavior between conditions, we found that children in the most concrete condition (IAV) interpreted behavior of both groups identically with $F(1,67) = 0.000, p = 1.000, \eta^2_p = .000$. In the adjective condition we found numerically more favorable interpretations of ambiguous behavior for the “abstract negative” group whereas in the noun condition positive interpretations of ambiguous behavior in the noun condition was numerically more favorable for the “abstract positive, yet differences in interpretations did the were not significant with $F(1,64) = 0.463, p = .499, \eta^2_p = .008$ for the adjective and $F(1,64) = 0.912, p = .343, \eta^2_p = .014$ for the noun condition so that we cannot interpret this result further. Figure 18 illustrates positive interpretations of ambiguous group behavior between conditions.

![Ambiguous Behavior Task](image)

**Figure 18.** *Averaged number of positive interpretations of ambiguous group behaviors in the three experimental conditions of Experiment 2.*

**Discussion**

In our second experiment, we had investigated whether and how language abstraction influenced formation of group representation when group behavior lacked a clear valence. We therefore designed a storybook presenting two novel social groups in six narratives each, half of
the narratives displaying positive, the other half negative behavior. We applied two linguistic manipulations, (1) manipulating which kind of behavior was described in abstract formulations, i.e. either the positive or negative behaviors, and (2) on which level of abstraction, i.e. IAVs, or adjectives, or nouns, the respective behavior was described. We expected children to evaluate groups more positively when positive behaviors were reported abstractly, and vice versa, evaluate groups more negatively when negative behaviors were reported abstractly. We moreover expected the strength of evaluation, thus the difference between group evaluations to increase with growing levels of abstraction of the descriptions.

However, contrary to our expectations we found no systematic effect of manipulation of language abstraction for none of the dependent variables, i.e. difference between group evaluation was not systematically more pronounced when we used higher levels of language description, neither did we found a consistent effect of language abstraction on group positivity or negativity, i.e. only in singular cases did the language pattern we used (abstract positive and concrete negative, respectively abstract negative and concrete positive) descriptions result in rather positive or rather negative representations of groups.

For group evaluation, we found a numerical differences, that was however not significant. Yet, it is at least notable that we found a significant difference for representations of both groups only in the condition for which we used nouns to describe group behavior. We might therefore speculate that the use of nouns could lead to an evaluative distinct representation of a rather positive (Group 1: “abstract positive”) and rather negative (Group 2: “abstract negative”) group assuming an effect of language abstraction only when descriptions are given in the most abstract form possible. However, since we did not find an interaction effect regarding language abstraction we do not have sufficient empirical data for this conclusion.
There are several plausible explanations for the pattern of results, specifically the lack of a systematic effect of language abstraction. Firstly, it is conceivable that in the absence of diagnostic group behavior, linguistic subtleties in behavioral descriptions alone are not sufficient to influence the formation of group representation. That is, when groups behave equally positive and negative language alone does not drive representations to be either positive or negative. It is however, at least noteworthy, that a distinct evaluation of both groups was found in the most abstract condition, only. This might be an indication that children are susceptible to language difference, yet in a different than we hypothesized i.e. children’s state of language comprehension might counteract the assumed systematic influence of language abstraction as hypothesized based on the LCM. In the discussion for Experiment 1, we already pointed out that adjectives play a subsidiary role in early language development (Rössl, 2007). It seems possible, that abstraction effects depend on the real-life relevance of linguistic categories in the target’s sample vocabulary. Likewise, the absence of linguistic effects could be explained through children’s unfamiliarity with certain linguistic categories or specific terms and words. Even though we included elementary school children in our sample that are likely to have a more extensive language repertoire compared to preschool children, it is still possible that age range of our sample was not broad enough and the vocabulary we used not appropriate for the intended purpose. With this pattern of results we cannot clearly decide whether children are simply not or less susceptible to language abstraction than hypothesized, whether language abstraction is no effective manipulation when information is completely ambiguous or whether we lack effects based on the developmental status of the sample.

Moreover, interpretation of the results is complicated because of the inaccuracy of our measurement scales, since reliability has proven to be rather poor. It is possible that the items we
used for measurement covered too diverse aspects of evaluative dimensions and children, thus, did not combine the items, e.g. helping, being nice, general liking, to coherent evaluative dimensions. Also, reliability might have been deficient due to the already mentioned differing relevance of linguistic categories.

For reward allocation, we found a marginally significant effect, indicating a larger allocation of rewards for the “abstract nice” group compared to the “abstract naught” group. Although rewards increased for the “abstract positive” group with increasing abstraction of descriptions while rewards for the “abstract negative” remained on a similar level and children seemed to more strongly reward the “abstract positive” group, (without sanctioning the “abstract negative” group), our analysis did not yield a qualifying effect of language abstraction. This reward distribution pattern might as well be influenced by children’s tendency for justice and equality despite contextual information (Sigelman & Waitzmann, 1991). We repeatedly noted children justifying an even allocation of rewards between the groups reasoning with fairness of allocation. Children furthermore issued allocation strategies not necessarily related to group evaluation. For example, some children awarded cones with less ice cream scoops arguing that too much ice cream would cause stomach pain. Decision for a lower number of rewards in this case does not reflect less favorable group evaluation, but rather a protective attitude. Another problem arises when gradients for rewards are not clearly quantifiable. For example, we chose four toys of different value, assuming that increasing value of toys correlates with increasing levels of attractiveness for children. It is however, possible that children awarded toys based on their personal preferences and that these preferences not necessarily correlated to the value of toys. Thus, children’s reward choice may not necessarily have coincided with the reward structure we thought of.
For new member evaluation the pattern of results seemed to partially fit our hypothesis, with more negative evaluations for the “abstract negative” group and marginally more positive evaluations for the “abstract positive” group, yet these differences were not significant, neither was there a significant difference regarding new member preference. There was moreover no effect of regarding levels of abstraction.

For the ambiguous behavior task we found a slightly higher number of positive interpretations for ambiguous behavior of the “abstract positive” group, yet again the difference in frequency of interpretation did not reach significance. Moreover, comparing the interpretation patterns between conditions demonstrated that there is no clear trend of interpreting ambiguous behavior. While on a numerical level differences between groups seemed to be absent in the IAV condition, interpretations were favorable for the “abstract negative” group in the adjective condition and favorable for the “abstract positive” group in the noun condition.

Reviewing the results of our second experiment, we have to conclude that linguistically manipulated information about groups did not systematically affect children’s formation of group attitudes. We neither found effects on children’s evaluation of new members, children’s preference decisions, neither on group nor on an individual level, or children’s interpretation of ambiguous behavior. For this experiment, we therefore conclude that linguistic manipulation, with few exceptions, does not seem to be strong enough to outweigh the effect of behavioral ambiguity regarding group evaluation and attitude formation.

We have further used this study to investigate how age-differences might influence the effect of linguistic manipulation. However, our results demonstrate no systematic age-related effects. The absence of an age-related effect could, on the one hand, be easily explained by the general absence of any effect of linguistic manipulation. That is, if we consider specifics in our
design, e.g. the complete ambiguity of group behavior, responsible for the absence of a linguistic effect, it is only logical that we could not measure age related differences. On the other hand, it is conceivable that children of both age groups were too young, respectively not yet susceptible to subtle linguistic manipulation and our manipulation could not unfold since we did not target the appropriate age group. Our data however does not allow clear conclusions on that topic.

**Linguistic Manipulation of Group Perception for positive or negative novel groups**

*(Experiment 3a)*

When designing the current research project, we focused on the following questions: How does language abstraction influence formation of intergroup attitudes when groups are presented with consistently good or bad behavior (Experiment 1), respectively when groups are presented ambiguously with equally good and bad behaviors (Experiment 2)?

For the third part of our research project, we aligned the experimental design more closely to the theoretical foundations of our hypothesis. For the previous experiments, we have relied on the novel group paradigm yet didn’t introduce novel groups in an intergroup context even though an intergroup context has been essential for demonstrations of the LIB. We argue that using an intra- instead of an intergroup context in the previous experiments could be a crucial factor for the presence or the absence of linguistically induced effects on group perception. The circumstance that group members only interacted with one another might, thus, have reduced the relevance of group behavior for participants, because group membership might

---

3 We preregistered our experiment with the Open Science Framework, the registration is retrievable under https://osf.io/ujg69/
have not been salient respectively relevant for information processing.

Hence, with the following experiment we explored whether manipulating the abstraction of behavioral descriptions affected children’s attitudes when (1) the valence of group was ambiguous, i.e. members of both groups displayed equally favorable and unfavorable behaviors and (2) all behaviors were presented in an intergroup context. While the main focus of our investigation remained on the formation of intergroup attitudes in early childhood (Experiment 3a), we included an adult comparison sample (Experiment 3b) to explore whether developmental differences might affect the formation of novel intergroup attitudes.

For that purpose, we used to different versions of a storybook in which members of two novel groups behaved equally often in favorable and unfavorable ways toward a member of the other group. We also varied the level of abstraction in the descriptions of the behavioral episodes as previously done in Experiment 2. That is, we described desirable behaviors with abstract terms and undesirable behaviors with concrete terms for one group (Group1: “abstract positive”); for the other group, desirable behaviors were described with concrete terms and undesirable behaviors with abstract terms (Group2: “abstract negative”).

Additionally, we included a control condition in which actions of both groups were described, using only concrete terms independent of the valence of the described behavior. Implementing this control condition should allow us to attain a baseline evaluation of both groups and thereby explore potential causes for evaluation biases in our material.

We investigated effects of linguistic manipulations using a number of child-oriented dependent measures as detailed below. We expected participants to make no evaluative differences between the two novel groups in the control condition. In the experimental conditions, we expected (a) children to evaluate Group 1 (“abstract positive”) more positively
than Group 2 (“abstract negative”) and further expect (b) these valence differences to be more pronounced the larger the contrast between concrete and abstract. Thus, the valence effect is expected to be absent in the DAV-versus-DAV condition, smallest in the DAV-versus-IAV condition and larger in the DAV-versus-Adjective condition and the DAV-versus-Noun condition.

**Method**

**Materials.**

*Storybook.* For structure and format of the storybooks, we relied on our designs of the previous experiments. Thus, storybooks initially introduced two novel groups, which were distinguished by group label (Lupis vs. Nifos) and T-shirt color (blue vs. red). In the following sequences, both groups were portrayed displaying an equal number of favorable and unfavorable behaviors. For all behavioral episodes, we established an intergroup context with a member of one group acting positively or negatively towards a member of the respective other group. Anew, sequences were presented on a double-page with the context established left and the behavior in question on the right. Pictures were sized 13 x 16 cm. A majority of the behavioral sequences we used in Experiment 2 (protecting, sharing, tidying up, being irresponsible, loud, aggression, cheering up, not sharing/being greedy, recklessness) was utilized in this experiment, as well. We further used two sequences portraying opposites of those used in Experiment 2 (discouraging, refusing to wash) and an additional sequence (helping behavior) not used before.

Within the behavioral narratives we used two linguistic manipulations. Firstly, we again manipulated which type of behavior was described abstractly, i.e. either the positive (Group 1) or the negative (Group 2) behavior. Secondly, whereas all narratives of the situational context were identically worded in DAVs across all conditions, descriptions of the focal actions varied in that they either contained a DAV (control condition), an IAV, a person-describing adjective, or a
person-describing noun.

Participants.

We conducted this experiment with 146 pre-school children and 157 elementary school children. For our pre-school sample we set an age limit excluding children younger than four years. Yet, some of the kindergarten teacher nevertheless included children aged three in the group of participating children (n = 8), whose data we excluded for analysis. Within this subsample we further excluded data from two participants due to erroneous conductance of experimental procedure (e.g. not reading all behavioral sequences), data from two participants due to massive disruptions while assessing the dependent variables (i.e. other children trying to clue the answers), two participants for which the experimenters reported low attention and, consequently, deemed results not interpretable as well as seven participants for which experimenters noticed language comprehension difficulties raising considerable doubts about the interpretability of their data. No children were excluded from the elementary school sample.

Thus, our final sample contained 125 pre-school children (\(M_{\text{Age}}= 4.73 \text{ years}, \ SD = .723, 59.2\% \text{ girls, } 40.8\% \text{ boys}\) and 157 first and second graders in elementary school (\(M_{\text{Age}}= 7.13 \), \(SD = .749, 51.6 \% \text{ girls, } 48.4\% \text{ boys}\).

A considerable number of children in our final sample identified as bilingual. Particularly, in the pre-school sample, 30 children reported growing up with a native language other than German (e.g., Albanian, Russian, Portuguese, Turkish, or Arabic), and 15 children declared to speak another language besides their native German (e.g. Turkish, Russian, English or Korean). In our elementary school sample five children natively spoke a different language from German (e.g. Japanese, Serbian or Russian), further 17 children commanded German and a second language as their mother tongue (e.g. Pashto, Turkish, Chinese, Finish or Farsi). For none of these children was there any indication of comprehension problems during the experimental
We derived the appropriate sample size from structurally similar studies investigating labeling effects (Degner, in prep.) and planned to include 30 participants per condition and sample. Distribution of participants over conditions can be seen in Table 6. Demographic variables of participants did not differ significantly between groups, neither concerning age, $F(3, 281) = 0.532, p = .660$ nor concerning gender $\chi^2(3) = 3.358, p = .340$.

<table>
<thead>
<tr>
<th></th>
<th>DAV</th>
<th>IAV</th>
<th>Adjective</th>
<th>Noun</th>
</tr>
</thead>
<tbody>
<tr>
<td>preschool sample</td>
<td>27</td>
<td>34</td>
<td>32</td>
<td>32</td>
</tr>
<tr>
<td>elementary school sample</td>
<td>38</td>
<td>39</td>
<td>43</td>
<td>37</td>
</tr>
<tr>
<td>Sample Total</td>
<td>65</td>
<td>74</td>
<td>75</td>
<td>69</td>
</tr>
</tbody>
</table>

Dependent Variables.

The assortment of dependent variables was largely adapted from Experiment 2. That is, for group evaluation, we individually presented the two group pictures and asked children to judge “How often do the Lupis/Nifos help other children?”, “How nice are the Lupis/Nios?”, respectively “How often do the Lupis/Nifos annoy other children?“ and “How mean are the Lupis/Nifos?”. We additionally asked children “How much do you like the Lupis/Nifos?”. Children judged each group on 3-point smiley scales and were given additional verbal quantifiers.

Furthermore, we applied a slightly altered version of the reward allocation measure.

When conducting Experiment 2, we repeatedly observed that children distributed the ice-cream
reward based on deliberating how ice-cream effects well-being (“too much ice-cream causes stomach pain”) or how much ice-cream they were allowed themselves. We subsequently feared that reward allocation for the ice-cream item was not associated to group evaluation but to unrelated arguments, thus, we replaced the ice-cream cones by plush animals. Moreover, we added the reward scale for each item by one, thereby letting participants choose between four instead of three rewards. By presenting four rewards, we intended to avoid tendencies for balancing decisions and selection of the scale mean. Accordingly, for the reward allocation measure participants could choose between four different kinds of rewards given to each group: the amount of candies (one, two, three or even four candies per child), the size of a present (a small, quite small, big or even a very big present), the kind of toy (a single crayon, a sandbox bucket and shovel, a pedal car or a playhouse) and finally the number of plush animals (only one, a giraffe, two, i.e. giraffe and a monkey, three, i.e. a giraffe, a monkey, and an elephant or even four, i.e. a giraffe, a monkey, an elephant, and a lion).

For the group interaction preference measure, both groups were presented side by side and participants were asked to select whether they would prefer to play with either of the groups, both groups or neither of the groups.

After completion of the group evaluation items, we introduced two new group members of the same gender as the participant. Congruent to the group evaluation measure for new member evaluation, we first asked children to judge “How often does this Lupi/Nifo help (other children)?”, “How nice is this Lupi/Nifo?”, respectively “How often does this Lupi/Nifo annoy (other children)?” and “How mean is this Lupi/Nifo?”. Additionally, we asked children “How much do you like this Lupi/Nifo?”. We used the same 3-point smiley and verbal quantifiers scales from the group evaluation measure.
For the new member interaction preference measure, children indicated who they would prefer to play with on the same response format as in the group measure.

The last part of the dependent variables set was constituted by an altered version of the ambiguous behavior task. Firstly, we limited pictures of ambiguous situations to four overall and, secondly, we presented actors without any signs of group membership, i.e. no labels are used to characterize the actor and all clothing is whitened. For each ambiguous situation, we presented a short behavioral description ending with an either positive or negative interpretation of the situation (e.g., insulting someone vs. sharing something with someone). Children were subsequently asked to decide to which group, i.e. the Lupis or Nifos, the actor more likely belongs to and in which color, i.e. red or blue, his or her clothing would have to be colored, accordingly.

Procedure.

Data collection for our kindergarten sample started on January 30th, 2017 and was completed on April 6th, 2017. Overall, eight kindergartens from the city center and suburbs of Hamburg took part in the experiment and nine experimenters (eight female, one male) were involved in gathering the data. For our elementary school sample, we collected data in three schools in the state of Schleswig-Holstein with the help of six experimenters (all female). The experimental procedure for both samples was congruent and corresponded with procedures of the previous experiments. That is, experimenters collected parental consent forms in kindergarten groups or classrooms and led children to a separate room, where they were tested individually. We secured that not more than four, in exceptional circumstances, five children were tested simultaneously in one room. Again, we split the experimental procedure, i.e. one of the experimenters assessed the demographic variables and read the storybook to the child, while the
second, condition-blind experimenter, assessed the dependent variables in above stated order. The distribution of conditions was randomized, i.e. the children themselves chose which of the externally identical looking storybooks they wanted to be read from. Randomization was again limited when documentation of the experiment revealed a large imbalance in condition distribution. Under this circumstance, we presented children a limited set of storybooks to choose from – of course only as long as balanced distribution was approximated.

Design.

Similar to Experiment 2, for which we applied a structurally identical linguistic manipulation, the design of this experiment follows a 2 (pattern of behavioral descriptions: positive behavior abstract + negative behavior concrete vs. negative behavior abstract + positive behavior concrete) within-subjects x 2 (evaluative dimension: positive vs. negative item valence) x 4 (level of abstraction: DAV vs. IAV vs. ADJ vs. Noun) between-subjects design.

The linguistic manipulation of the within-factor aimed at creating an opposed evaluation of groups, i.e. forming a positive representation of the group by describing positive behavior abstractly and forming a rather negative representation of the group by describing negative behaviors abstractly. To prevent potential demand effects, we counterbalanced between participants which group label – Lupis versus Nifos – was assigned to Group 1 and Group 2.

Including the between-factor allowed us to implement a second linguistic manipulation by varying the level of abstraction in descriptions. Once again, variations of levels of abstraction were only applied to the specific behaviors, i.e. contextual depictions were always given in the most concrete level (DAV). Behavioral descriptions were, thus, either given in DAVs for the control condition or in the experimental conditions in IAVs person-describing adjectives or person describing nouns only. With a four factor between participant manipulation and a
counterbalanced two factor within-manipulation, this design produced eight different storybook versions.

Hypothesis.

We expected a significant effect of group membership as a result of manipulating which type of group behavior was described abstractly (within-factor). Although both groups were presented with an equal amount of positive and negative behavior, we expected our linguistic manipulation to outweigh the behavioral ambiguity for both groups. That is, we expected evaluations of Group 1 (for which positive behaviors were described abstractly and negative behaviors concretely, i.e. “abstract positive”) to be more positive compared to Group 2 (for which negative behaviors were described abstractly and positive behaviors concretely, i.e. “abstract negative”). We upheld this assumption, despite yielding no significant effects of group membership in Experiment 2 for which we applied a structurally similar linguistic manipulation. We argue that effects of linguistic manipulation are more likely to unfold in an intergroup context as presented in Experiment 3 compared to the previously presented intragroup contexts.

For the control condition, we however expected no difference in group’s evaluations. This hypothesis extended to all dependent variables, i.e. we expected children to allocate more rewards to Group 1 (“abstract positive” group), prefer to interact with Group 1 (“abstract positive” group) and more likely ascribe positive behaviors to Group 1 (“abstract positive” group). We further expected these group evaluations to transfer to new member evaluations, thus expected new members of Group 1 (“abstract positive” group) to be evaluated more positively and be preferred interaction partners. However, as new member evaluations were derived from group level evaluations, we expected these evaluations to be less strong.

Our second hypothesis referred to the manipulation of the between-factor. For the
manipulation of the between-factor, in general, we expected that the evaluative differences between groups would be more pronounced the higher the level of language abstraction used for the behavioral descriptions. That is, we expected no significant group differences in the control condition, but expected the evaluative difference between groups to increase with increasing levels of abstraction. We extended this hypothesis to all dependent variables, expecting reward allocation, group as well as new member preferences, new member evaluation and ascription of positive behaviors to more strongly favor Group 1 (“abstract positive” group) in the more abstract conditions.

Results

All statistical analyses were run with IBM® SPSS® Statistics (Version 23) for Mac. Results for each of the dependent variables are reported independently. Analysis in general includes both samples, if sample related age effects occur, these will be reported separately. Parallelly, all analyses were run in the excluded sample of three years old.

Group evaluation.

To determine whether linguistic abstraction influenced group evaluation, we conducted a 2 (pattern of behavioral descriptions: “abstract positive” vs. “abstract negative”) x 2 (evaluative dimension: positive evaluation vs. negative evaluation) x 4 (level of abstraction: DAV vs. IAV vs. ADJ vs. Noun) repeated measures ANOVA that, however, showed no effect of linguistic abstraction, $F(3,278) = 1.194, p = .312, \eta_p^2 = 0.013$.

A 2 (pattern of behavioral descriptions: positive behavior abstract + negative behavior concrete vs. negative behavior abstract + positive behavior concrete) x 2 (evaluative dimension: positive vs. negative item valence) repeated measures ANOVA yielded a marginally significant
interaction of group valence and item valence with $F(1,281) = 3.520, p = .062, \eta^2_P = .012$.

Average evaluations of groups were numerically higher for Group 1 (“abstract positive” group) with $M = 2.30$ ($SD = .46$) compared to Group 2 (“abstract negative” group) with $M = 2.25$ ($SD = .45$). Negative evaluations were numerically higher for Group 2 (“abstract negative” group) with $M = 2.06$ ($SD = .52$) compared to Group 1 (“abstract positive” group) with $M = 1.99$ ($SD = .54$).

Further analyzing potential interactions between group evaluation and evaluative dimension by running a $2 \times 2$ (pattern of behavioral descriptions: positive behavior abstract + negative behavior concrete vs. negative behavior abstract + positive behavior concrete) x 2 (evaluative dimension: positive evaluation vs. negative evaluation) repeated measures ANOVA separately for each condition, demonstrated that – opposed to our hypothesis – the interaction of group and item valence was significant only in the most concrete condition, $F(1,64) = 4.619, p = .035, \eta^2_P = .067$ yet neither for linguistic manipulation with IAVs with $F(1,72) = 0.211, p = .647, \eta^2_P = .003$ nor adjectives with $F(1,74) = 0.555, p = .459, \eta^2_P = .007$ nor nouns with $F(1,68) = 2.089, p = .153, \eta^2_P = .030$
DEVELOPMENTAL PERSPECTIVE ON LIB

"abstract positive" group vs. "abstract negative" group:

- **DAV**
  - Positive evaluations: ~2.5
  - Negative evaluations: ~1.5

- **IAV**
  - Positive evaluations: ~2.0
  - Negative evaluations: ~2.0

- **ADJ**
  - Positive evaluations: ~2.0
  - Negative evaluations: ~2.0
Comparing evaluations between there was no significant effect of sample. Table 7 contains the results for the remaining sample comparisons.

**Table 7. Interaction between group and item valence for each condition and sample combination of Experiment 3a.**

<table>
<thead>
<tr>
<th>Sample</th>
<th>DAV</th>
<th>IAV</th>
<th>Adjective</th>
<th>Noun</th>
</tr>
</thead>
<tbody>
<tr>
<td>preschool</td>
<td>$F(1,26) = 1.354, p = .255, \eta^2_p = .050.$</td>
<td>$F(1,33) = 5.750, p = .022, \eta^2_p = .148.$</td>
<td>$F(1,33) = .045,$</td>
<td>$F(1,31) = .761, p = .390,$</td>
</tr>
<tr>
<td>elementary</td>
<td>$F(1,37) = 6.027, p = .019, \eta^2_p = .140$</td>
<td>$F(1,38) = 2.124, p = .153, \eta^2_p = .053$</td>
<td>$F(1,42) = 1.144, p = .291,$</td>
<td>$F(1,36) = 1.336, p = .255,$</td>
</tr>
</tbody>
</table>

Group evaluation was measured with three items for the positive and two items for the negative dimension with the different items appraising groups on varying levels of abstraction.

For analysis of group evaluation, we averaged results on both evaluative scales. Reliability for all scales was overall poor, with likewise weak reliability in both subsamples. Reliability scales are
summarized in Table 8.

Table 8. Reliability of the positive and negative evaluation scales for group evaluation in Experiment 3a.

<table>
<thead>
<tr>
<th></th>
<th>Group 1 (“nice group”)</th>
<th>Group 2 (“naughty group”)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Positive Evaluation</td>
<td>Negative Evaluation</td>
</tr>
<tr>
<td>Pre-School</td>
<td>α = .566</td>
<td>α = .587</td>
</tr>
<tr>
<td>Elementary School</td>
<td>α = .599</td>
<td>α = .591</td>
</tr>
</tbody>
</table>

Group preference.

Interaction preferences were not influenced by linguistic manipulation, $\chi^2(9) = 3.385, p = .947$ nor did preference decisions differ significantly between sample groups, $\chi^2(3) = 1.401, p = .705$. Overall, 281 participants gave an interaction preference. Half of the participants (n= 141, 50.2%) preferred an interaction with the “abstract positive” group, 23 participants expressed a preference for both while eight participants expressed no preference for any of the groups.

Distribution of interaction preferences between conditions is depicted in Figure 20.

![Figure 20](Image)

Figure 20. Preference for groups in absolute numbers for the four experimental conditions of Experiment 3a.
Reward Allocation.

Reward Allocation was applied as a second and more indirect measure of group evaluation, for which children distributed four different types of rewards to both of the groups.

Conducting a 2 (reward allocation: “abstract positive” vs. “abstract negative”) x 4 (level of abstraction: DAV vs. IAV vs. ADJ vs. Noun) repeated measures ANOVA, we found a marginally significant effect of group valence, $F(1, 278) = 3.443, p = .065, \eta^2_{p} = .012$, yet no further qualification of reward distribution by linguistic manipulation, $F(3, 278) = .258, p = .836, \eta^2_{p} = .003$.

The pattern of reward distribution was consistent with our hypothesis, i.e. more rewards were allocated to the “abstract positive” group ($M = 2.95, SD = .81$) compared to the “abstract negative” ($M = 2.84, SD = .81$) Differences in reward allocation between groups did approach significance only in the most concrete condition using DAVs with $F(1, 64) = 2.722, p = .104, \eta^2_{p} = .041$. Figure 21 shows reward allocation between conditions.

![Reward Allocation](image)

*Figure 21. Distribution of rewards in the four experimental conditions of Experiment 3a.*

For the purpose of exploring potential effects of sample group we conducted a 2 (reward
allocation: “abstract positive” vs. “abstract negative”) x 2 (sample group: pre-school vs. elementary school) x 4 (level of abstraction: DAV vs. IAV vs. ADJ vs. Noun) ANOVA. Neither the interaction between sample and group valence, $F(31,274) = .069, p = .793, \eta^2_P = .000$, nor three-way interaction between sample, group level and abstraction of linguistic manipulation, $F(3,274) = 1.537, p = .205, \eta^2_P = .017$, gave indication for an effect of sample.

Table 9. Main effect of group membership for reward allocation in all experimental conditions of Experiment 3a

<table>
<thead>
<tr>
<th></th>
<th>DAV</th>
<th>IAV</th>
<th>ADJ</th>
<th>Noun</th>
</tr>
</thead>
<tbody>
<tr>
<td>$F$</td>
<td>$F(1,64) = 2.722$,</td>
<td>$F(1,72) = .050$,</td>
<td>$F(1,74) = .700$,</td>
<td>$F(1,68) = 1.142$,</td>
</tr>
<tr>
<td>$p$</td>
<td>.104,</td>
<td>.824,</td>
<td>.406,</td>
<td>.289,</td>
</tr>
<tr>
<td>$\eta^2_P$</td>
<td>.041</td>
<td>.001</td>
<td>.009</td>
<td>.017</td>
</tr>
</tbody>
</table>

Reliability of reward scales was satisfactory for both groups, with $\alpha = .753$ for the nice group reward scale, respectively $\alpha = .736$, for the naughty group reward scale.

New Member Evaluation.

We conducted a 2 (group membership: “abstract positive” group vs. “abstract negative” group) x 2 (evaluative dimension: positive evaluation vs. negative evaluation) x 4 (level of abstraction: DAV vs. IAV vs. ADJ vs. Noun) repeated measures ANOVA which yielded no significant effect by level of abstraction, $F(3,278) = 0.411, p = .745, \eta^2_P = .004$, yet a marginally significant interaction of group membership and item valence, $F(1,278) = 3.689, p = .056, \eta^2_P = .013$. Consistent with our hypothesis, we found that new members of the “abstract positive” group were evaluated more positively ($M = 2.45, SD = .51$) compared to new members of the “abstract negative” group ($M = 2.41, SD = .53$); this pattern was reversed for negative evaluations with new members of the “abstract negative” group being evaluated more negatively ($M = 1.81, SD = .66$) compared to new members of the “abstract positive” group ($M = 1.73, SD = .64$).
Again, when comparing results in the four linguistic conditions, this interaction of group and item valence was nonsignificant in all conditions with $F(1, 68) = 1.249, p = .268, \eta^2_p = .018$ for nouns, with $F(1, 74) = .140, p = .710, \eta^2_p = .002$ for adjectives, with $F(1, 74) = 0.140, p = .538 \eta^2_p = .005$ for IAVs and approached significance only in the most concrete condition, with $F(1, 64) = 2.461, p = .109 \eta^2_p = .040$. Figure 22 contains an overview on group evaluations between conditions.
DEVELOPMENTAL PERSPECTIVE ON LIB

**DAV**

- Positive evaluations
- Negative evaluations

**IAV**

- Positive evaluations
- Negative evaluations

**ADJ**

- Positive evaluations
- Negative evaluations
Exploration for effects of sample produced two significant results. The interaction of item valence and sample was significant within the adjective, $F(1,73) = 4.342, p = .041$, $\eta^2_P = .056$, and the noun condition, $F(1,67) = 4.541, p = .037$, $\eta^2_P = .063$.

Within the adjective condition, elementary school children evaluated the “abstract negative” group more positively compared to the “abstract positive”, while pre-school children were evaluated the “abstract nice group” more positively. Within the noun conditions, evaluations of the sample groups diverged significantly for negative evaluations of the “abstract negative” groups (pre-school: $M = 2.16, SD = .77$ vs. elementary school: $M = 1.64, SD = .57$), while all other evaluations were similar in direction.

Reliability of scales was acceptable ($\alpha \geq .710$) for negative evaluations of both groups and slightly weaker for positive evaluations ($\alpha \geq .660$). By tendency, reliability seems slightly higher in the pre-school sample. Table 10 contains an overview about reliability scores in both samples.

Figure 22. New member evaluation for both groups in all experimental conditions of Experiment 3a.
Table 10. Reliability of the positive and negative evaluation scales for new member evaluation in Experiment 3a.

<table>
<thead>
<tr>
<th>Group 1 (&quot;nice group&quot;)</th>
<th>Group 2 (&quot;naughty group&quot;)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Positive Evaluation</td>
<td>Positive Evaluation</td>
</tr>
<tr>
<td>Pre-School α = .691</td>
<td>α = .705</td>
</tr>
<tr>
<td>Elementary School α = .609</td>
<td>α = .694</td>
</tr>
<tr>
<td>Negative Evaluation</td>
<td>Negative Evaluation</td>
</tr>
<tr>
<td>Pre-School α = .726</td>
<td>α = .872</td>
</tr>
<tr>
<td>Elementary School α = .626</td>
<td>α = .638</td>
</tr>
</tbody>
</table>

New Member Preference.

Interaction preferences were not qualified by level of abstraction in the narratives, $\chi^2(6) = 3.659, p = .723$, nor did interaction preferences differ between sample groups, $\chi^2(2) = .021, p = .990$. From overall 282 participants 132 (46.2%) preferred an interaction with a new member of the nice group, 30 participants expressed no preference for any of the both new members. Distribution of interaction preferences between conditions is depicted in Figure 23.
Ambiguous Behavior Task.

In our variation of the ambiguous behavior task, we created behavioral episodes in which group membership of the actor was uncertain, i.e. the actor was not labeled, neither did clothing color reveal group membership. Children had to subsequently ascribe the behavior to either a member of the nice or a member of the naughty group and indicated group membership by giving the group a label and the appropriate clothing color. For analysis of the ambiguous behavior task, we first determined the number of responses that were consistent with our linguistic manipulation of group valence and for that purpose calculated how often children ascribed positive behavior to the nice group and negative behavior to the naughty group. We further conducted our analysis in three steps. That is, we firstly analyzed responses for indication of group membership by label only, secondly, we analyzed responses for indication of group membership by naming the color of clothing and finally, we analyzed responses only when indication of group membership by label and color matched, i.e. indicated the same group.

For responses indicated by label only, we conducted a repeated measures ANOVA that yielded no effect of abstraction, $F(3,278) = .999, p = .394, \eta^2_p = .011$.

On average children ascribed 1.95 out of four behaviors ($SD = .94$) in conformance with the valence of group membership, with $M = 1.01$ (out of two behaviors, $SD = .78$) for negative behaviors ascribed to the “abstract negative” group and $M = .93$ (out of two behaviors, $SD = .73$) for ascription of positive behavior to the “abstract positive” group. The most consistent ascriptions of group membership to behavior were given in the most concrete condition ($M = 2.15, SD = 1.14$); the least consistent ascriptions in the most abstract condition ($M = 1.84, SD = 1.08$), however the effect of linguistic manipulation was not significant therefore we cannot interpret these differences any further. There was moreover no effect for sample, $F(1,280) = 1.889, p = .170, \eta^2_p = .007$. 
In a second step, we investigated indications of group membership by naming of clothing color. When analyzing the number of consistent interpretations between conditions, we found no significant effect of linguistic manipulation, \( F(3,278) = 1.269, p = .285, \eta^2_P = .014 \) either.

Children on average ascribed 1.93 out of four behaviors (SD = 1.24) in conformance with the valence of group membership with \( M = 1.0 \) (SD = .82) consistent ascriptions of negative behaviors to the abstract negative group) and \( M = .93 \) (SD = 0.72) ascriptions of positive behaviors to the abstract positive group.

The highest number of consistent ascriptions was given by participants in the most concrete condition (\( M = 2.14, SD = 1.28 \)) yet since there is no effect of linguistic manipulation we cannot further interpret numeric differences in consistent ascription between conditions.

There was no effect of sample again, \( F(1,280) =2.037, p = .155, \eta^2_P = .007 \).

In a third step, we analyzed only those responses of participants whose indication of label and color of clothing implied the same group. For example, if a participant ascribed positive behavior to a member of the Lupis we analyzed the response only if the participant as well identified the appropriate group color, i.e. blue for a member of the Lupis.

With this pre-selection we had to exclude responses of overall 85 participants, for which the stated pairing of group label and group clothing color were inconsistent.

Analysis for differences in ambiguous behavior interpretation between conditions yielded no significant effects of linguistic manipulation, \( F(3,193) = 1.348, p = .260, \eta^2_P = .021 \).

The number of consistent interpretations was \( M = 1.65 \) (SD = .75) overall, \( M = 0.96 \) (SD = .69) for ascriptions of positive behavior to the “abstract positive” group, respectively \( M = 0.69 \) (SD = .51) for ascriptions of negative behavior to the “abstract negative” group.

The highest number of consistent interpretations was given by participants in the
adjective condition ($M = 1.79, SD = .81$), the second highest number of consistent interpretations was given in the most concrete DAV condition ($M = 1.73, SD = .79$) – yet since there was again no effect of linguistic manipulation, we cannot further interpret these results.

Differing from the previous analysis, we found a marginally significant interaction of sample, $F(1,195) = 3.547, p = .061, \eta^2_P = .018$. Comparing response patterns between conditions demonstrated that the difference in consistency for interpretations is largest when IAVs were used for linguistic manipulation, with elementary school children being more likely to group label and color consistently than pre-school children; this difference is marginally significant $F(1,50) = 3.846, p = .055, \eta^2_P = .071$. Figure 24 illustrates the number of consistent ascriptions for the four experimental conditions.
Discussion

We conducted our third experiment to further explore how representations of novel groups are formed when behavioral information about respective novel groups is ambiguous. In contrast to the design of Experiment 2, we presented groups in the experiment at hand in an intergroup context and investigated whether linguistic abstraction would influence impression formation under these circumstances.

For group evaluation we found no favoritism for the “abstract positive” group over the “abstract negative” group, a favorable group evaluation of the “abstract positive” respectively an unfavorable evaluation of the “abstract negative” group was only found in the most concrete condition, i.e. the condition we basically used as a control condition with no linguistic manipulations of behavior descriptions. There was, moreover, no clear preference for the “abstract positive” over the “abstract negative” group. Moreover, quite opposed to our hypothesis, we found that evaluations were most distinct in the most concrete condition rather

![Figure 124. Consistent ascriptions of behavior for a) indication of group membership by label only, b) indication of group membership by clothing color only and c) indication of group membership by matching color and label.](image)
than in the most abstract condition. Even though the effect of linguistic manipulation was not significant it is however remarkable that these results did not hint towards an effect of language abstraction as hypothesized.

The result pattern for new member evaluations was similar, i.e. members of the “abstract positive” group were evaluated more positively and vice versa; members of the “abstract negative” group were evaluated more negatively, with this evaluative difference between groups being marginally significant. There was no further qualification by the abstraction of narratives, yet again, numerically group differences were most pronounced in the most concrete condition. Analysis of preferences did not show a clear preference for new members of either group.

Our analysis of the ambiguous behavior task demonstrated that children on average managed to interpret just under half of the behavioral episodes consistent with group valence. With this seemingly random result pattern, it seems very well possible that children’s behavioral ascriptions were not influenced by our first linguistic manipulation of group positivity, but might stem from the ambiguity of group behavior. That is, with groups behaving equally positively and negatively, participants might not perceive consistent behavior for either of the groups and therefore demonstrated an indecisive respectively random response pattern. Linguistic abstraction as our second manipulation, again, asserted no further influence on interpretation of ambiguous scenes. Moreover, it is remarkable, that a quite large number of participants, i.e. 35%, were not able to appropriately match group labels to clothing color. This result might be an indication for the instability of built group representations. If children were not able to consistently identify groups, it is at least possible that results on all dependent measure did not accurately reflect evaluations of the designated group.

In summary, we can conclude that manipulating which type of behavior was described
abstractly, respectively concretely, did not systematically nor consistently influence whether positive or negative evaluations of groups were formed. Moreover, contrary to our hypothesis, we found the tendency that group evaluations were more pronounced in our “control condition” – even though we have to regard these results with caution since an effect of linguistic manipulation did not yield significance.

We have already speculated that this deviation from our hypothesis might be grounded in age related differences of language comprehension. First of all, it is conceivable that children are simply not as susceptible to linguistic manipulation or that children are susceptible to other linguistic categories than hypothesized based on the taxonomy of the LCM. Secondly, we might have used vocabulary that wasn’t appropriate, e.g. because children were not familiar with the words we used to describe the focal action. Thus, unfamiliarity might have cancelled out potential effects of language abstraction. Thirdly, it is still also possible that children’s still developing ability of language processing has not yet allowed linguistic manipulation to come into effect.

We have moreover to consider that the marginally significant difference between groups we reported were not based on our linguistic manipulation either but could be a random effect caused from imbalances in the stimulus materials.

Conclusively, our data cannot confirm our hypothesis regarding the effects of language on impression formation. We find no systematic evidence that the level of abstraction used for behavioral descriptions influenced group representation or subsequent evaluations. We have speculated that not necessarily the inexistence of language effects on impression formation was the reason for our results, but considered possible age-related sample effects hindered our manipulation. To explore whether language abstraction differently affects attitude formation in a
child or adult. we conducted a subsequent experiment with an adult’s sample.

**Linguistic Manipulation of Group Perception for positive or negative novel groups with adults (Experiment 3b)**

Repeating our third experiment with adults followed the purpose to explore age effects and compare the influence of linguistic manipulation on attitude formation in early childhood to adulthood. To allow a proper comparison between sample groups, we adapted experimental proceedings as closely as possible. Therefore, for the description of the experiment’s method we will largely refer to previously described proceedings and only highlight necessary required changes.

**Method**

**Materials.**

The materials we employed for the current experiment were identically to the materials used in Experiment 3a, i.e. we adapted pictures and narratives unaltered. However, as we conducted the adult version of the experiment in form of an online study, we digitalized the paper versions of the storybook. We therefore transformed the storybook print templates to GIF-Files suitable for online presentation. Structure of the storybook and order of narratives remained the same as in Experiment 3a. To allow presentation of the storybooks on different display formats, we fixed the size for narratives to 666 x 444 pixels.

**Participants.**

Overall, 265 students followed the study advertisement and called up the starting page of our online survey. 26 students tried to access the survey by smartphone or tablet, an access route
we had excluded because the presentation of storybooks was impaired for mobile devices. Hence, these 26 participants were excluded from participation. Further 74 students failed the attention check at the beginning of the survey and were excluded from participation for that reason.

Language comprehension was an essential prerequisite for study participation, we therefore decided to only include data of participants whose native language was German or German among another language. Thus, before initiating data analysis, we excluded eight participants who did not meet the required language background.

To successfully assess the potential effects of our manipulation, it was equally essential to our study that participants paid careful attention to pictures and corresponding narratives about the Lupis and Nifos. We therefore excluded an additional five participants, who reported technical problems while retrieving the online version of the storybook. That does not include participants who reported prolonged loading times, but those indicating that one or more pictures were not displayed at all and one participant, who indicated technical problems without further specification.

To ensure that participants could concentrate sufficiently on the presented narratives, we moreover excluded those participants who indicated comprehension difficulties due to the shortness of the given time frame. Altogether 14 participants reported that the twelve second time frame, in which narratives were presented, was too short to fully read or comprehend or accurately remember the information and were subsequently excluded from the analysis.

Finally, we excluded an additional two participants due to incomplete response sheets, i.e. participants who responded to less than 50 % of the dependent variables.

Thus, our final sample consisted of 137 participants, who were almost evenly distributed
between the four conditions with 36 participants in the DAV condition, 35 in the IAV condition, 32 in the adjective condition and 34 in the noun condition.

The participants’ age ranged notably between 17 and 48 years with $M = 24.69 \ (SD = 6.37)$. Corresponding with gender distribution in psychological study programs, the majority of our student sample was female (81%). Due to our study criteria, all participants had native speaker competence of German, 22 participants additional reported a second mother language such as English, Turkish or Chinese. Unsurprisingly, the vast majority of participants indicated their course of studies as main occupation (94.9%), five participants reported a current employment as main occupation, one participant self-identified as currently unemployed while another one identified as self-employed.

Dependent Variables.

With our adult sample we also collected data on group evaluation, reward allocation, group preferences, new member evaluation and preference and measured interpretation of ambiguous behavior. While we used the pictorial and verbal stimuli for the questions identical to the ones used in the previous experiment, we modified the response scales for group and new member evaluation. By presenting five instead of three possible responses we allowed a more differentiated assessment of both groups. For questions assessing behavioral typicality, i.e. “How often do the Lupis/Nifos help other children?” the offered response options were “never” (German: “nie”), “seldom” (German: “seltan”), “sometimes” (German: “manchmal”), “often” (German: “oft”) and “always” (German: “immer”). For questions regarding evaluating traits, i.e. “How nice respectively naughty are the Lupis/Nifos?”, and assessing the liking, i.e. “How much do you like the Lupis/Nifos?”, we offered the response options “not at all” (German: “gar nicht”), “little” (German: “ein bisschen”), “somewhat” (German: “somewhat”), “quite
much” (German: “ziemlich”) and “very much” (German: “sehr”).

We kept the smiley faces accompanying the verbal responses scales, but did not add pictures for the additional responses so that smiley faces marked the lowest, highest and mid-scale point.

Due to erroneous programming of the survey, response options for preference decisions were presented forced-choice. That is, participants had to issue their preference for either of the groups or new individuals but could not express preference for both or neither.

Procedure.

We advertised participation on the online platform SONA. SONA is run by the University of Hamburg and allows associated researchers to advertise lab as well as online studies. The majority of potential participants on this platform are students from the University of Hamburg, especially the psychological department, because these students have to gather a certain amount of participation credit as part of their bachelor studies. Occasionally, students from other universities, who do not maintain an own platform for study advertisement, participate as well. Our experiment was advertised with the title “Once upon a time… - a children’s book project”. Potential participants were simply told that they would read a children’s book and afterwards would be asked some questions about the book. Interested participants were forwarded to the survey platform Qualtrics by hyperlink, on which our experiment was hosted.

On the introductory page of the experiment, participants were once more informed that they would read a children’s book and afterwards had to answer questions about characters and narratives, thus they were cautioned to read carefully. Participants were further informed about the anonymity of data collection, voluntariness of participation and the estimated length of the
experiment, which was specified with approximately 20 to 30 minutes.

With linguistic manipulation as the core of the experiment, we considered close attention to the narratives an essential requirement for participation. We therefore employed an attention check on the second page of the experiment. For this attention check, we wrote a paragraph that supposedly informed participants about the need to collect a minimum of person-related data. We ended the paragraph with a reference to anonymity of data collection and analysis. At the bottom of the paragraph participants were in bold letters prompted to enter their age. However, in the middle of the paragraph between information on person-related data and anonymity, we instructed participants about the necessity to pay close attention during the experiment and revealed that contrary to instructions in bold at the bottom of the paragraph, it was required to enter the present year (2017) instead of the participants’ age. Participants failing to pass this attention check were forwarded to the end of the survey.

We then assessed demographic variables, such as age, sex, current occupation and native language. Participants who self-identified as non-native German speakers and classified their level below fluent were forwarded to the end of the survey as well.

On the following page participants were once more reminded to pay close attention to the following narratives. They were further instructed, that pages of the storybook would turn automatically (after 12 seconds) and to start the study as soon as they are ready. Participants were then randomly assigned to one of the eight conditions.

After reading the stories, participants were guided through the assessment of the dependent variables, presented in the established order, i.e. group evaluation, group preference, reward allocation, new member evaluation, new member preference and the ambiguous behavior task.
Participants were finally asked whether they had technical difficulties while working on the experiment, were requested to specify the difficulties in case they indicated some, could give a general feedback to the survey and were finally forwarded to SONA where they automatically received participation credits.

Design.

The design of the experiment was held equal to the children’s version of the experiment. The design thus followed a 2 (pattern of behavioral descriptions: “abstract positive” vs. “abstract negative”) within-subjects x 2 (evaluative dimension: positive vs. negative item valence) x 4 (level of abstraction: DAV vs. IAV vs. ADJ vs. Noun) between-subjects design.

We again counterbalanced whether the group label Nifo or Lupi was assigned to Group 1 respectively Group 2 and thus created eight experimental conditions.

Hypothesis.

We expected our adult participants in general to acquire evaluative representations of both groups, i.e. to evaluate the “abstract positive” group, for which we described positive behavior abstractly, more positively and less negatively than the “abstract negative” group, for which we described negative behavior abstractly. However, as we described all behaviors for both groups only concretely in our control condition, we expected no difference between groups in the respective condition.

We further expected evaluative differences between groups to be more pronounced the more abstract behavioral description of positive, respectively negative, behaviors were. In these predictions we included all of the employed dependent variables. Yet, as previous research with adults has demonstrated that nouns only partially fit into the predictions of the LCM we could
not be entirely sure if positive and negative evaluations would rise lineally over all conditions.

Results

Again, all statistical analyses were run with IBM© SPSS© Statistics (Version 23) for Mac. Results for each of the dependent variables are reported independently.

Group evaluation.

Due to a coding error, the first 59 participants were presented with an incorrect version of the dependent variables. Precisely, for the positive evaluation of the Lupis participants were first correctly presented with the item “How often do the Lupis help other children?” and the corresponding response scale, yet instead presenting “How much do you like the Lupis?” with its corresponding scale secondly, participants were presented with a repetition of the first item “How often do Lupis help other children?” combined with the scale for the second item, i.e. the assessment of liking the Lupis. Although this error seems quite severe, it seemingly affected results only marginally.

Thus, we decided to present results for the data as collected. However, to allow comparisons with results when accounted for the error we excluded the “How nice (...)?” item for the Lupis and conducted a 2 (description of group behavior) x 2(item valence) x 4 (language abstraction) repeated measure ANOVA and yielded a comparable result, i.e. $F(3,133) = 10.222, p = 0.000, \eta^2_p = .187$. Similarly, the interaction of group and item valence was significant only in the DAV condition, $F(1,35) = 4.838, p = .035, \eta^2_p = .121$ and in the adjective condition, $F(1,31) = 16.829, p = .000, \eta^2_p = .352$. Thus, the evaluation patterns within conditions are consistent with the patterns reported for analysis for the complete data set. Results were further unaffected if the item in question was only excluded for participants presented with the erroneous coding but retained for all participants with correct versions. That is, the 2 (description of group behavior) x 2(item valence) x 4 (language abstraction) repeated measure ANOVA yielded an almost identical result, i.e. $F(3,133) = 9.868, p = 0.000, \eta^2_p = .182$. Results for evaluative patterns within conditions are similar as well with significant interactions only in the DAV $F(1,35) = 4.486, p = .041, \eta^2_p = .114$, and the adjective condition, $F(1,31) = 16.213, p = .000, \eta^2_p = .343$. 


We conducted a 2 (pattern of behavioral descriptions: “positive abstract” vs. “negative abstract”) x 2 (evaluative dimension: positive vs. negative item valence) x 4 (language abstraction: DAV vs. IAV vs. ADJ vs. Noun) repeated measures ANOVA. The results demonstrated a significant interaction between group valence, item valence and abstraction of narratives, $F(3,133) = 9.924, p < .001, \eta^2_p = .183$.

Consistent with our assumptions, adult participants evaluated in general the “abstract positive” group more positively ($M = 3.32, SD = .06$) compared to the “abstract negative” group ($M = 3.22, SD = .06$), and evaluated the “abstract negative” group more negatively ($M = 2.97, SD = .06$) compared to the “abstract positive” group ($M = 2.88, SD = .06$).

In the DAV as well as in the IAV condition we found the “abstract negative” group being evaluated more positively and less negatively than the “abstract positive”, in the adjective group the “abstract positive” group being evaluated more positively and less negatively compared to the “abstract negative” group and for the noun group evaluations of the “abstract negative” group more negatively yet also more positively than the “abstract positive” group. This difference in evaluation was however only significant for the DAV, $F(1,35) = 4.702, p = .037, \eta^2_p = .118$, and adjective condition, $F(1,31) = 16.009, p = .000, \eta^2_p = .034$. An overview of group evaluation can be found in Figure 25.
DEVELOPMENTAL PERSPECTIVE ON LIB.

The diagrams show the comparison between "abstract positive" and "abstract negative" groups in terms of positive and negative evaluations, measured on a scale from 1 to 4. The graphs illustrate that both groups receive positive and negative evaluations, with the "abstract positive" group generally receiving slightly higher positive evaluations compared to the "abstract negative" group. The bars indicate the mean evaluations with error bars showing the standard deviation.
For analyzing the differences in group evaluation, we summarized the positive, respectively negative, evaluative items for both groups, which resulted in four evaluative scales. Reliability for all scales ranged between minimal $\alpha = .771$ for negative evaluations of the naughty group and $\alpha = .849$ for positive evaluations of the naughty group and, hence, was overall acceptable.

Group preference.

Interaction preferences were not systematically influenced by linguistic manipulation, $\chi^2(3) = 2.931 \ p = .404$. Due to an error in coding, participants could only indicate preference for either of the group over the respective other, yet were not presented with the option of preferring both or none of the groups. Facing this decision of preference, 73 of 137 participants (53.3 %) indicated a preference for the “abstract positive” group over the “abstract negative” group. Figure 26 shows distribution of interaction preferences between conditions.
Reward Allocation.

A 2 (pattern of behavioral descriptions: “abstract positive” vs. “abstract negative”) x 4 (language abstraction: DAV vs. IAV vs. ADJ vs. Noun) repeated measures ANOVA yielded a significant interaction effect, $F(1,133) = 2.978, p = .034, \eta^2_P = .06$.

Analyzing reward allocation between conditions, demonstrated that for both rather concrete conditions participants allocated slightly more rewards to the “abstract negative” group than to the “abstract positive” group. This pattern was reversed in the more abstract conditions. However, favoritism of the “abstract positive” group regarding reward allocation approached the level of significance in the adjective condition only, $F(1,31) = 3.965, p = .055, \eta^2_P = .113$ yet neither for DAV with $F(1,35) = 1.471, p = .233, \eta_2P = .040$, nor IAV with $F(1,34) = .384, p = .539, \eta_2P = .011$ nor nouns, with $F(1,33) = .017, p = .897, \eta_2 = .001$. Figure 27 illustrates reward allocation in the four experimental conditions.

On average participants awarded $M = 2.5 \ (SD = .60)$ rewards to the “abstract positive” group and $M = 2.49 \ (SD = .58)$ to the “abstract negative” group. This difference in reward allocation was not significant, $F(1,136) = 0.180, p = .627, \eta^2_P = .001$. 

Figure 26. Group preferences between conditions.
For measurement of reward we used four item reward scales, which demonstrated sufficient reliability with $\alpha = .670$ for the “abstract positive” group reward scale, respectively $\alpha = .636$ for the “abstract negative” group reward scale.

New Member Evaluation.

A 2 (pattern of behavioral descriptions: “abstract positive” vs. “abstract negative”) x 2 (evaluative dimension: positive vs. negative item valence) x 4 (language abstraction: DAV vs. IAV vs. ADJ vs. Noun) repeated measures ANOVA repeated measures ANOVA yielded no significant interaction effect, $F(3,133) = 1.729, p = .164, \eta^2_p = .038$.

A 2 (pattern of behavioral descriptions: “abstract positive” vs. “abstract negative”) x 2 (evaluative dimension: positive vs. negative item valence) repeated measures ANOVA demonstrated a significant main effect of group valence, $F(1,136) = 5.299, p = .023, \eta^2_p = .038$, indicating that new members of both groups were evaluated differently. We moreover found a – theoretically less important – main effect of item valence $F(1,136) = 192.918, p = .000, \eta^2_p = .587$, indicating that positive and negative evaluation scores differed significantly from one to
Contrary to our hypothesis, evaluation of new group members was more positive \( (M = 3.51, SD = .048) \) for the “abstract negative” compared to the “abstract positive” \( (M = .3.49, SD = .046) \), although the difference is numerically quite small. Yet, consistent with our hypothesis, evaluations were likewise more negative for the “abstract negative” group \( (M = 2.59, SD = .054) \) in comparison to the “abstract positive” group \( (M = 2.25, SD = .051) \).

Even though we cannot compare evaluations between conditions for lack of significance regarding the effect of linguistic manipulation, it is at least noteworthy that only in the adjective condition we found a significant difference for predicted evaluation pattern, i.e. with more positive and less negative evaluations for a new member of the “abstract positive” group compared to a new member of the “abstract negative” group \( F(1,31) = 5.478, p = .026, \eta^2_p = .150 \), whereas interaction of group and item valence was not significant in the DAV (“control”) condition with \( F(1,35) = 0.468, p = .498, \eta^2_p = .013 \), nor in the IAV condition with \( F(1,34) = 0.304, p = .585 \eta^2_p = .009 \) nor in the noun condition with \( F(1,33) = 0.000, p = 1.000, \eta^2_p = .000 \). Figure 28 gives an overview on differences in new member evaluation.

For evaluation of new members, participants were again presented with the positive and negative evaluation scales for both groups. Scale reliability was between \( \alpha = .727 \) and \( \alpha = .780 \) and thus acceptable for all scales.
“Abstract positive” group
“Abstract negative” group

**DAV**

**IAV**

**ADJ**

- Positive evaluations
- Negative evaluations
New Member Preference.

New member preference was qualified by language abstraction, $\chi^2(3) = 1.324, p = .724$. Two participants did not respond to the preference item. Out of the remaining 135 participants, only a slight majority (52.6 %) preferred to interact with the “abstract positive” rather than the “abstract negative” group. For the IAV and the noun condition, choice of preference is an interaction with “abstract positive” group. However, for the DAV and the adjective condition, this pattern reverses with a preference for the “abstract negative” group.

Figure 28. *New member evaluation in the four experimental conditions of Experiment 3b.*

![Graph showing new member evaluation in the four experimental conditions of Experiment 3b.](chart1.png)

Figure 29. *Preference for new members in absolute numbers for the four experimental conditions of Experiment 3b.*

![Bar chart showing preference for new members in absolute numbers for the four experimental conditions of Experiment 3b.](chart2.png)
Ambiguous Behavior Task.

As a final measure we again used the ambiguous behavior task for which we asked participants to indicate whether a certain positive or negative behavior was conducted either by a member of the “abstract positive” group or the “abstract negative” group. Participants indicated group membership by giving the respective group label and the corresponding clothing color of the group. To analyze ascription of behavior we calculated how often participants ascribed the behavior as expected, i.e. positive behavior to the “abstract positive” group and negative behavior to the “abstract negative” group. However, we again split the analysis regarding the mode of group identification. That is, we firstly conducted the analysis for ascriptions by label only; secondly, for ascriptions by t-shirt color only. In a third step, we limited the analysis to those ascriptions for which label and t-shirt color correspondingly indicated the correct group membership.

For analysis of ascriptions by label only the influence of linguistic abstraction on the consistency of behavior ascription was significant, $F(3,133) = 4.081, p = .008, \eta^2_p = .084$

On average $M = 1.02 (SD = .73)$ positive behaviors were ascribed to the “abstract positive” group, respectively $M = 0.96 (SD = .82)$ were ascribed to the “abstract negative” group. Consistency of behavior ascriptions increased from the DAV to the IAV, peaked in the adjective condition, but considerably decreased in the noun condition.

In a second step, we analyzed ascriptions of behavior by clothing color and found a significant effect of language abstraction with $F(3,133) = 5.812, p = .001, \eta^2_p = .116$. On average participants ascribed $M = 1.03 (SD = .06)$ positive behaviors to the “abstract positive” group and $M = 0.92 (SD = .07)$ negative behaviors to the “abstract negative” group. Again, behavioral consistency increased from the DAV to the IAV condition and was largest for the adjective condition with a drop of consistency in the noun condition.
For groups correctly identified by label and group color we found a significant effect of language abstraction, $F(3,104) = 2.846, p = .041, \eta^2_p = .076$.

Participants on average ascribed $M = 0.96$ $(SD = .06)$ positive behaviors to the “abstract positive” and $M = 0.90$ $(SD = .07)$ negative behaviors to the “abstract negative” group. Pattern of ascription consistency in the four conditions matched the above-described trend. Figure 30 presents differences of positive interpretations of ambiguous behavior for all conditions.

The number of group misidentifications, i.e. cases in which participants indicated group membership contradictorily by using the correct group label but the wrong clothing color or vice versa, is inversed to the pattern of consistency ascriptions. That is, in the DAV condition participants more often gave contradictory indications of group membership (38.9 %), followed by the IAV (20 %) and noun condition (17.6 %), while the least amount of contradictory group identification was given in the adjective condition (6.3 %).
**Discussion**

This experiment was an almost unaltered rerun of our third experiment. We adapted an online appropriate version of our children’s book and presented it to an adult sample. Our goal with this experiment was to understand whether adults’ evaluations of the presented groups were similar or different from children’s. We assumed that this comparison would allow us to draw more accurate conclusions on the scope of language effects including sample variations regarding language comprehension and cognitive resources.

---

**Figure 30.** Consistent ascriptions of behavior for a) indication of group membership by label only, b) indication of group membership by clothing color only and c) indication of group membership by matching color and label in Experiment 3b.
We expected adults to easily comprehend and process the narratives and expected adult’s evaluations to be influenced by the level of language abstraction used for behavioral descriptions. More precisely, we expected adults to develop similar evaluations of both groups in the DAV conditions, in which all behaviors were described concretely. However, we predicted adults to develop a more positive and less negative evaluation of the “abstract positive” compared to the “abstract negative” group in all other conditions and the difference between group evaluations to increase with increasing abstraction of descriptions.

Yet, group evaluation patterns did not meet our expectations. We found a significant difference in group evaluation only for the DAV and adjective condition respectively, yet only for the adjective condition we found more favorable evaluations of the “abstract positive” group.

We observed a similar pattern of results for reward allocation, for which significant difference in reward allocation were only reported in the adjective condition.

Participants tended to favor an interaction with the “abstract positive” over the “abstract negative” group – yet there was no qualification by language abstraction.

Evaluation of new members tended to be more positive and more negative for the “abstract negative” group, however this difference was not further qualified by linguistic manipulation.

Overall, participants preferred an interaction with a new member of the “abstract positive” over the “abstract negative” group and interaction preferences were qualified by language abstraction.

The ambiguous behavior task demonstrated that participants were likely to attribute more positive behaviors to the “abstract positive” and more negative behaviors to the “abstract negative” group. This pattern of ascription was furthermore qualified by language abstraction.
Results of our fourth experiment focusing on linguistic biases was the first study we conducted with adults. Using an adult sample was an attempt to understand whether limited language or cognitive capacities in younger children could explain the lack of systematic language abstraction effects in our results. We predicted that adults would easily process the presented narratives and demonstrate lineally increasing favorability of the “abstract positive” over the “abstract negative” group with increasing abstraction in linguistic manipulation. The results of our experiment, however, did not deliver clear evidence for linear influences of language abstraction on impression formation.

We could, however, demonstrate that at least adjectives affected impression formation in the predicted way. Participants who were presented narratives using adjectives formed significantly more positive and less negative evaluations of the “abstract positive” group, although there was no actual difference in the valence of group behavior. This favorability of the “abstract positive” group was reproduced in the pattern of reward allocation and evaluation of new members. Participants moreover seemed to form clear expectations of groups’ behavior according to their perceptions of these groups as demonstrated in the ambiguous behavior task. Moreover, these favorable impressions of the nice group were reflected in preference decisions for groups, however not in the preference decision for new group members. Thus, it seems that using adjectives for behavioral descriptions might shape expectations about these groups’ behavior.

For our control condition in which we described all behaviors by using only DAVs we found at least partial favorability of the “abstract negative” group, even though representations of both groups were expected to be alike. This unexpected effect might be rooted in imbalances of the stimulus material. For example, it is possible that positive behaviors with which the “abstract
negative” group was presented seemed nicer to begin with, e.g. sharing is evaluated more positively than washing up. That is, even without verbal description of any linguistic kind there was an unintended imbalance in behavioral valence. Thus, it is possible that adults in the DAV condition reacted to imbalance of behavioral valence in the materials, but not to language form of any kind. Assuming such an imbalance in the stimulus material gives even more meaning to the results we demonstrated in the adjective condition—suggesting that the effect of adjectives on impression formation can be quite strong. Since, – expect for the ambiguous behavior task – the results are quite similar in the DAV and in the IAV condition, we presume that the verbal manipulation by IAVs was not able to overcome the potential bias induced by choice of stimulus material either.

Another explanation for the lacking effect of IAVs is, that language abstraction on impression formation does not operate in the assumed linear mode but rather be observed only when the level of abstraction reaches a certain threshold. Given the fact that nouns are on the upper range of language abstraction, it should be assumed that nouns would overcome this potential threshold and, thus, at least influence perceptions in a way similar to adjectives, if not excelling the scope of adjective-induced effects. However, in our experiment the use of nouns did neither influence impressions in a comparable manner to adjectives nor excelled the effect of adjectives on impression formation. The lacking effect of nouns might however be rooted in word choice. We used nouns from which we believed to be understandable for children and likewise appropriate descriptions of the presented behaviors. These nouns could however seem odd or unfamiliar when used in real life contexts. This might have reduced fluency of linguistic processing and counteracted a potential effect of abstraction. Future research, thus, should consider to investigate potential effects of nouns in impression formation with the use of more
adult-appropriate nouns or nouns that are considered abusive language, which is regularly used to refer to unpopular outgroups and minorities.

Taken together, our fourth study did not reveal the effects we presumed would be caused by linguistic manipulation, i.e. we could not demonstrate that language abstraction systematically influenced attitude formation. The results, however, do not rule out that language plays a role in impression and subsequent attitude formation. On the contrary, we were able to illustrate that verbal descriptions given in adjectives produced a rather positive or negative evaluation of depicted groups, although group behavior had no clear positive or negative valence. Thus, our study is a crucial first step in demonstrating how particular use of language can influence attitude formation.

We have, however, to take into account, that our choice of stimulus material might have affected results in two different manners. Firstly, we cannot, to this point, determine whether our failure to demonstrate a systematic effect of language abstraction illustrated the limits of linguistic influences on impression formation or stems from an erroneous selection of stimulus materials. Resolving this question and determining whether there is a systematic and linear effect of language on impression formation, must be subject to subsequent studies for which we recommend to ensure an equally positively and negatively rated set of behavioral sequences. Secondly, although we conducted this study with adults and explicitly asked for their opinions on both groups, participants were well aware that they evaluated protagonists of a children’s book. It is therefore possible, that participants’ responses did not reflect their opinions unbiasedly, but included considerations about potential reactions of children towards this material. If adults in the adjective condition assumed a better understanding of the story by children and thus a stronger reaction to the narratives, these presumptions might have affected participants’
responses. This reservation can, however, be easily revoked if future experiments customize more adult relevant stimulus material.

**General Discussion**

Research on language use in social context has demonstrated how people tend to use biased language in intergroup contexts. This phenomenon is known under the term Linguistic Intergroup Bias (Maass et al., 1989 and following for example Rubini & Semin, 1994; Maass et al., 1995; Franco & Maass, 1996). Biased on ample research evidence of the LIB in intergroup contexts it has been hypothesized that this language pattern can as well shape attitudes towards groups (Maass, 1999). As of yet, it remained however unclear, whether language abstraction is a consequence or a cause of biased intergroup attitudes. Thus, the aim of our research project was to advance the theory evolving around the LIB by investigating how language abstraction influences impression formation towards novel groups.

Considering previous research, we can rely on the knowledge that intergroup attitudes are developed from a very young age on (Bigler, 1997; Nesdale, 2001, Bigler & Liben, 2006, Raabe & Beelmann, 2011). We therefore decided to investigate attitude formation in a respectively young age group and conducted our experiments with kindergarten and elementary school children. For this purpose, we designed a set of three sequential experiments, exploring how language abstraction affects different parts of impression formation in children. We subsequently developed an adult version of our third experiment to allow comparisons between language effects in children’s and adults’ impression formation.

Our studies have rendered a variety of results that may help to investigate the influence of language abstraction on impression formation further.
Firstly, children seem to quickly acquire attitudes towards novel social groups based on information about group behaviors. More precisely, a unique presentation of a storybook, i.e. verbal information accompanied by pictorial stimuli, was sufficient to create either positive or negative attitudes towards novel groups. However, this observation is limited to contexts in which groups are presented behaving unanimously positive or negative. That is, a prerequisite for children to acquire clearly positive or negative attitudes towards novel groups seems to be a strong contingency between group membership and a certain type of behavior.

Secondly, when children are presented with unanimous information about novel groups, the abstraction in which the verbal information is given does not influence children’s perceptions of groups. If information creates a unanimous association between a group and a certain evaluation, it seems that the content of information overrules potential effects of informational linguistic form, i.e. the abstraction in which information is given.

Thirdly, the abstraction of information did not systematically affect group evaluation neither for groups behaving unanimously positive or negative, respectively, (clear behavioral valence) nor for groups behaving equally positive and negative (ambiguous behavioral valence). Instead, our results were rather inconclusive with significant results for different types of linguistic manipulation (e.g. either DAVs or adjectives).

Fourthly, children demonstrated consistency in their evaluations of groups, i.e. the result pattern we observed for group evaluation was repeated in children’s choice of reward allocation and to some extent also transferred to new member evaluation.

Fifthly, when we presented adults with the storybook displaying ambiguous intergroup behavior, we found that adults tended to give the most pronounced positive evaluations for the
“abstract positive” group, respectively negative evaluations for the “abstract negative” group, in the adjective condition.

Even though our experiments do not support our hypothesis of language abstraction systematically influencing attitude formation in childhood, our results to some extent allow the conclusion that communication of intergroup attitudes could still be important regarding attitude formation. We observed that, if children are given unanimous behavioral information about novel groups, in either a positive or a negative way, they will develop positive or negative attitudes towards these groups, accordingly. That is, from a young age on, children seem to be sensitive to the content of information they receive and actively use this information for navigation in the social world. Thus, a balanced, fair and unbiased communication about social group from parents, guardians and educators can possibly contribute to the formation of positive and egalitarian attitudes during childhood.

Our results, however, do not deliver unequivocal evidence on the role of language abstraction on impression formation in early childhood. Notwithstanding, the synopsis of our four experiments allows some conclusions on the question if and how language abstraction might influence attitude formation at all.

We found that, if the content of information itself paints a straightforward picture about groups (unambiguous group behavior in Experiment 1), the mode of information transmission, i.e. language abstraction, might be either too subtle to assert an additional influence on impression formation or counteracts the impression formation based on informational content. We neither found a systematic effect of language abstraction on impression formation the content of information was ambiguous regarding group positivity and negativity, i.e. groups behaved equally positive and negative. It seems however possible that in other ratios of
ambiguity, e.g. behavioral episodes containing a 60-40 ratio of positive to negative information about groups, an effect of language abstraction might come to effect. Regarding the absence of a systematic effect of language abstraction it is moreover conceivable that children did not grasp verbally conveyed concepts of group negativity or positivity but used own judgements of the behavior favorability as the baseline for group evaluations. For example, washing up after playing outside might be a behavior evaluated positively by adults but neutral to irritating by children.

Our results, so far, do not provide strong evidence that linguistic abstractness influences group impression formation in young children, rather did we report rather inconsistent outcomes of our experiments. For example, we in general expected evaluations of groups to be more pronounced when children were targeted with abstractly worded stories. In our second experiment, results fitted these assumptions, with the most pronounced evaluations observed in the noun condition. Though, the results of our third experiment proved to be exactly the opposite with children indicating the strongest evaluations in the most concrete condition. For our experiments with the adult sample, we however observed, that significant evaluative differences in the adjective condition as hypothesized, i.e. participants evaluated the “abstract positive” group significantly more positively and less negatively compared to the “abstract negative” group. It is remarkable that this response pattern was reflected in nearly all dependent variables. Considering this contradictory pattern of results and the fact that quite some of the results only barely reached levels of significance, we can only cautiously conclude that at least the use of some linguistic categories might affect impression formation – even if not in the developmental stage that we assumed but rather later in adolescence or adulthood. For future research on language abstraction with children it might be recommendable to use pre-tested idioms and
expressions for usage frequencies and participants’ familiarity with them, for future research
with adults it might be recommendable to test the replicability of our results regarding the effect
of adjectives on attitude formation in adults.

Given the replicability of the effect we found regarding adjectives affect on impression
formation, this finding has severe implications - all the more if we consider that this effect was
obtained after a one-time exposure to a very limited set of information. We could therefore
assume that the linguistic form of media reports (for research on the LIB regarding media reports
see e.g., Dragojevic, Sink & Mastro, 2016; Geschke, Sassenberg, Ruhrmann & Sommer, 2010;
for influences of television-media reports on linguistic biases in viewers see Gorham, 2006) can
shape perceptions of novel groups, especially when information is given repeatedly or
conversation about specific social groups is an ongoing process. Thus, at least in partial there
might be a contribution of biased language to perpetuation of biased intergroup attitudes as
assumed by Maass (1999). Further research on this topic might be an important step to expand
theoretical foundations of the LIB.

Still, we have to address the question why the (limited) effect of language abstraction
found for adults is absent in our research with young children. In our search for explanations, we
found a variety of reasons that might be accountable for our findings. First of all, we assume that
age-related difficulties in categorization processes might be responsible for the absence of a
consistent language effect in our children sample. In the ambiguous behavior task participants
were asked to ascribe a certain behavior to either of the groups. Indication of group membership
was given by assigning the correct label and the t-shirt color of the respective group. We
originally intended this task to provide insights on whether children’s evaluations would transfer
to expectations about future group behaviors. The responses revealed, however, that children had
major difficulties to correctly match visual and verbal clues indicating group membership. Therefore, it is possible that children didn’t build appropriate representations of groups. For future research we therefore recommend to enhance frequency of presentation or other experimental designs to ensure that proper categorization of groups is learnt before testing further effects of language abstraction.

In this regard, another point worth considering is whether a combination of verbal and visual clues is the best mode to establish group typicality and group differentiation. Research on social categorization by Baron et al. (2014) has already demonstrated that the impact of verbal clues by far outperforms the impact of visual clues in the matter of categorization. For future research it could, thus, be beneficial to forego establishment of visual distinctions and rely on the effect of verbal labels alone.

Another explanation for the diverging results in both age groups might, moreover, be an age-related difference in language processing. Especially young children are still in the process of mastering language comprehension and production. Although we tried to reduce complexity of the narratives to an age-appropriate level, it is within the realm of possibilities that children’s language abilities are not yet differentiated enough to allow subtle manipulations of language abstraction to take effect. We know for example, from research on psycholinguistics, that children learn nouns earlier respectively more easily than adjectives (e.g. Goldin-Meadow, Seligman & Gelman, 1976; Mintz & Gleitman, 2002) or verbs (McDonough, C., Song, L., Hirsh-Pasek, K., Golinkoff, R. M., & Lannon, R. (2011). It is therefore conceivable that the assumed effect of abstraction in general, or the assumed linearity of the effect specifically, does not apply to children. For example, concrete terms inhere the additional advantage to feature lesser complexity, so that children might process concrete descriptions better and faster than other
linguistic terms. It is moreover possible, that the vocabulary we used deviated from the vocabulary children used in everyday interaction. Sandhofer and Smith (2007) demonstrated that experimental manipulations in laboratory settings did not match child-parent-interaction in real life and thus laboratory findings might mismatch observations in real-life-contexts. This might have been true for our word choice in our experiments as well. That is, language comprehension and processing as well as language familiarity might have also impaired the effectiveness of our manipulation. Considering the lack of an effect in the age range of our sample, it might be possible that until the age of ten, children are not influenced by language subtleties. Since our adult participants, for whom we could demonstrate partial effects of language abstraction, were on average 24 years old; it is worth considering that language influences attitude formation only in later childhood or early adolescence. Drawing on results of previous research allows an approximation towards the age range that might be interesting for future research. For example, Salès-Wuillemin et al. (2014) have demonstrated that the production of the LIB is provable in eleven-year-old teenagers. Although, the production of a LIB is a different phenomenon than the susceptibility to language abstraction regarding impression formation, both require a certain level of language mastery. In that case, we would assume that linguistic manipulation of impressions should be effective in a similar age. Thus, if future studies aim to determine the age from which on children are susceptible to linguistic variation, we would recommend a sample age range of 11 to 14 years.

Another limitation of our study is frequency of stimulus presentation. Our experiment was designed as a “one-time-exposure” to both groups, yet learning of social groups in real-life-context will happen gradually, over time and with repeated exposure. This assumption does not necessarily contradict the results of our first experiment for which we found distinct evaluations
for both groups after a single exposure, since for this experiment we presented groups with a clear positive and negative valence of behavior. It is rather an explanation for lack of language abstraction when group valence was presented ambiguously as done in Experiment 2 and 3, which might be similar to presentation of social groups in real-life-contexts. For future research is should thus be considered that language abstraction comes to effect only over time and therefore experiments should be designed to present stimulus sets more frequently or over longer research periods.

Inconsistency of results might have also been caused by stimulus materials. Although, we tried to choose our storybook narratives carefully, we did so without pretesting them for positivity and negativity. So, we did not know for certain, whether positivity and negativity of behavior was balanced for groups in each stimulus set neither whether positivity and negativity of stimulus sets were comparable between conditions. It is therefore possible that our results were at least partially produced by reactions towards the inward positivity or negativity of the stimulus sets rather than to the verbal information given with the stimulus sets. Especially, the results of our third experiment indicate such a bias. We would have expected no difference between group evaluations in our control condition, i.e. the condition in which all behaviors for both groups were described in DAVs, yet children’s evaluations of both groups diverged. Therefore, it is likely, that the stimulus set contained behaviors that were more positive for one group than the other and that this bias is reflected in children’s evaluations. We suggest that the results of our Experiment 3b (“the adult sample”) support this assumption. Similar to children’s responses, adults demonstrated a slight bias for the “abstract negative” over the “abstract positive” group in the rather concrete conditions. This unexpected direction of evaluation might be due to biases in the stimulus material. If there was an imbalance of valence in our stimulus
sets causing the specific result pattern we observed, then it is possible that the distortion in our materials might have prevented us from finding a larger range of effects. Accordingly, the fact that we found an effect for adjectives despite a potential bias in the materials needs emphasis. We might therefore assume that the effect of adjectives on impression formation might be quite strong. Based on our experiences, we recommend future researchers to carefully pretest materials and balance positivity and negativity in the presented narratives.

Despite a careful selection of stimulus material, research with children requires a likewise careful conception of dependent variables. From our experience, children reacted frustrated and bored to repetitious presented variables. For example, in our first experiment, we asked “How many of the Lupis / Nifos are...”? and offered children six evaluative dimensions, which made them answer structurally similar questions about twelve times in a row. Assessing this variable, we observed that children tended to become less attentive to the experiment and less interested in participating, which might impair data quality. Children moreover used response strategies that do not match our predictions and could undermine our interpretation of the results. For example, we expected for allocation of rewards the “abstract positive” groups to be awarded with the highest number of rewards. We cannot however be certain that our evaluation of the highest reward was identical with what children assumed to be the highest or best reward. When we presented a reward scale with toys, we arranged toys in the order of monetary value. We thereby coded the most valuable toy as the most desirable and expected it to be chosen as the highest reward. Children, however, might have followed their own preferences when allocating rewards, e.g. allocating their favorite toy and not the most valuable toy to a group might – from children’s point of view – be a better reward than allocating the monetary most valuable toy. Under such circumstances, our analysis of reward allocation would not accurately reflect children’s
intentions. Children demonstrated moreover other lines of reasoning that could as well have counteracted our hypothesis of reward allocation. For example, children distributed smaller amounts of sweets or ice cream, because they assumed that larger portions would cause stomach ache and wanted not to cause pain to either of the groups or simply followed normative demands regarding what they know is allowed when it comes to consumption of sweets. These are further illustrations why children’s data in our experiments should be interpreted with caution.

Our experiments were designed to demonstrate empirical evidence for the effect of language abstraction, more specifically linguistic biases in social communication, on impression formation in adulthood. Our results do so far however not support this hypothesis. We found no systematic effect of language abstraction on impression formation in childhood nor in adulthood. We could however demonstrate how the use of adjectives affected impression formation of adulthood, so we suggest that it is possible that language abstraction at least to some intend affects impression formation. We presented a number of limitations from our studies that might have prevented us from finding the systematic effect we expected. Despite our lack of results, we still consider it within the realm of possibilities, that language abstraction does effect impression formation when a certain developmental stage is reached, even though this effect might not be linear and similar to attitude formation in real-life-context might take place over time only.

Conclusion

Without a doubt, language is a major means of human communication. Language is used to convey appeal as well as request, information as well as emotion. It has been suggested that linguistic biases in language might function not only as a reflection of existing intergroup attitudes but might as well assert causal influence on group evaluation and thereby transmit
(biased) intergroup attitudes. Our research mainly focused on formation of group representation in children. While we found that children could quickly acquire group evaluations of unambiguously presented groups in a novel group paradigm, acquired group evaluations were not influenced by linguistic manipulation. When children were presented with ambiguously behaving groups linguistic manipulation of behavioral descriptions did not yield in clearly favorable evaluations of groups for which we used abstract positive descriptions neither clearly unfavorable evaluations of groups for which we used abstract negative descriptions. We found however an effect of adjectives influencing group evaluation in adults. With this overall pattern of results, we have to concede that we found little to none effect of systematic effects of linguistic manipulation on impression formation and based on our experiments cannot support the hypothesis that linguistic biases play a significant role in intergroup attitude formation and transmission. Future research might however help to explore whether there are effects of language abstraction other than hypothesized.
REFERENCES


Communication (pp. 313-330). New York: Psychology Press


**APPENDIX A: EXAMPLE OF A STORYBOOK (PICTURES AND NARRATIVES)**

Manipulation of behavioral narratives: blue = descriptive verb; green = interpretative verb; red = adjective.

All usage rights are reserved for the Arbeitsbereich Sozialpsychologie, Universität Hamburg. Pictorial stimuli are scaled to minimum quality in order to prevent unauthorized usage. Material can of course be shared upon request. In case of interest please contact Prof. Dr. Juliane Degner via juliane.degner@uni-hamburg.de.


Und der Lupi hat seine Jacke und seine Schuhe schon an. /

Und der Lupi hat sich schon ganz alleine fertig gemacht. /

Der Lupi ist schon richtig selbstständig.
So, kann es losgehen, raus zum Spielen. |
| --- | --- |
| Und hier auf dem Bild ist eine Nifo *(zeigen)*. Die Nifo und die anderen Kinder laufen richtig schnell. Oh je, dieses Kind ist hingefallen *(zeigen)*. Es hat sich am Knie weh getan. Das tut ein bisschen weh. Das Kind weint auch ein bisschen. | Und guck mal was hier los ist.

Die Nifo *(zeigen)* läuft an dem Kind vorbei und dann lacht die Nifo wirklich doll. /

Die Nifo hänselt das andere Kind. /

Die Nifo ist wirklich ganz schön gehässig.
Oh weh. Das Kind, das sich weh getan hat, bleibt traurig am Boden sitzen. |
| Und hier siehst du wieder einen Lupi *(zeigen)*. Der Lupi spielt zusammen mit einem anderen Kind Ball. Jetzt ist der Ball an den Zaun gerollt. Und da am Zaun, da steht ein großer Hund *(zeigen)*. Oh je, wie das Kind guckt *(zeigen)*. Das Kind hat Angst vor Hunden und erschreckt sich wirklich doll. | Guck mal, der Lupi!

Der Lupi stellt sich vor das andere Kind und ruft so laut, dass der Hund wegläuft. /

Der Lupi beschützt das andere Kind. /

Der Lupi ist wirklich mutig.
Puh, ein Glück! Da ist das andere Kind aber froh.
Jetzt können die beiden weiterspielen. |
Hier spielen einige Kinder im Sand und eine Nifo ist auch dabei (zeigen). Im Sandkasten kann man prima spielen und eine Sandburg bauen. Dieses Kind hier ist noch ganz neu in der KiTa (zeigen). Es guckt der Nifo zu. Eigentlich möchte das neue Kind auch gerne eine Schaufel haben.

Im Sandkasten sind noch andere Kinder und auch ein Lupi (zeigen). Die Kinder wollen im Sand mit Murmeln spielen. Und der Lupi, der hat ein großes Netz, da sind alle Murmeln drin (zeigen).


Hier spielt eine Lupi zusammen mit anderen Kindern fangen (ticken). Da hinten ist die Lupi (zeigen). Dieses Kind hier (zeigen) ist so schnell gelaufen, da ist es hingefallen. Und es hat sich ein bisschen weh getan, sogar die Hose ist kaputt gegangen (zeigen) und das Kind weint auch ein bisschen.

Aber guck hier, die Nifo (zeigen).

Schau dir den Lupi mal genau an.

Jetzt regnet es schon wirklich doll. Alle Kinder gehen rein (zeigen), aber guck mal der Nifo hier (zeigen).
Der Nifo bleibt mit seinem Ball im Regen und lacht. / Der Nifo gehorcht einfach nicht. / Der Nifo ist ganz schön ungezogen.

Aber guck mal hier, die Lupi (zeigen).

Aber guck mal hier! Wie das aussieht. Alle Spielsachen, der Ball und die Eimer und das Seil und das Auto, liegen noch draußen rum und der Nifo (zeigen)?
Der Nifo geht einfach rein und lässt alles liegen. / Der Nifo räumt gar nicht auf. / Der Nifo ist wirklich faul.

Jetzt haben alle Kinder schon viel gespielt.


Guck mal hier, die Nifo (zeigen).
Iih, die Nifo hat noch Schmutz im Gesicht und an den Händen. /
Iih, die Nifo hat sich gar nicht sauber gemacht. /
Iih, die Nifo ist ganz schmutzig.
Da wird beim Abtrocknen sogar das Handtuch schmutzig (zeigen).

Jetzt ist die Lupi im Zimmer (zeigen). Aber die Lupi läuft nur noch auf Zehenspitzen. /
Aber die Lupi nimmt extra Rücksicht. /
Aber die Lupi ist wirklich rücksichtsvoll.
So wird keines von den anderen Kindern wieder wach.
APPENDIX B: EXAMPLE OF AN AMBIGUOUS BEHAVIOR TASK

<table>
<thead>
<tr>
<th>VP - NR</th>
<th>Datum:</th>
<th>VL:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Gruppenevaluation</td>
<td></td>
</tr>
<tr>
<td>1. Wie lieb sind die Lupis?</td>
<td>Gar nicht (1)</td>
<td>Ein bisschen (2)</td>
</tr>
<tr>
<td>2. Wie gemein sind Lupis?</td>
<td>Gar nicht (1)</td>
<td>Ein bisschen (2)</td>
</tr>
<tr>
<td>3. Wie lieb sind die Niños?</td>
<td>Gar nicht (1)</td>
<td>Ein bisschen (2)</td>
</tr>
<tr>
<td>4. Wie gemein sind die Niños?</td>
<td>Gar nicht (1)</td>
<td>Ein bisschen doll (2)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Gruppenhomogenität</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wie viele von den....</td>
</tr>
<tr>
<td>5. Lupis sind lieb?</td>
</tr>
<tr>
<td>6. Lupis sind nett?</td>
</tr>
<tr>
<td>7. Lupis sind fröhlich?</td>
</tr>
<tr>
<td>8. Lupis sind gemein?</td>
</tr>
<tr>
<td>9. Lupis sind böse?</td>
</tr>
<tr>
<td>10. Lupis sind unfreundlich?</td>
</tr>
<tr>
<td>11. Niños sind lieb?</td>
</tr>
<tr>
<td>12. Niños sind nett?</td>
</tr>
<tr>
<td>13. Niños sind fröhlich?</td>
</tr>
<tr>
<td>14. Niños sind gemein?</td>
</tr>
<tr>
<td>15. Niños sind böse?</td>
</tr>
<tr>
<td>16. Niños sind unfreundlich?</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Gruppenpräferenz</th>
</tr>
</thead>
<tbody>
<tr>
<td>17. Mit wem würdest du lieber spielen? Den...</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Einzelevaluation:</th>
</tr>
</thead>
<tbody>
<tr>
<td>18. Wie lieb ist der neue Lupi?</td>
</tr>
<tr>
<td>19. Wie gemein ist der neue Lupi?</td>
</tr>
<tr>
<td>20. Wie lieb ist der neue Niño?</td>
</tr>
<tr>
<td>21. Wie gemein ist der neue Niño?</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Individualpräferenz?</th>
</tr>
</thead>
<tbody>
<tr>
<td>22. Mit wem würdest du lieber spielen? Dem ...</td>
</tr>
</tbody>
</table>
Ankreuzen, ob sich das Kind für **die erste** oder **die zweite** Verhaltensoption entscheidet!

<table>
<thead>
<tr>
<th></th>
<th>Option 1</th>
<th>Option 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>23.</td>
<td><img src="image1.png" alt="Image" /></td>
<td><img src="image2.png" alt="Image" /></td>
</tr>
<tr>
<td>24.</td>
<td><img src="image3.png" alt="Image" /></td>
<td><img src="image4.png" alt="Image" /></td>
</tr>
<tr>
<td>25.</td>
<td><img src="image5.png" alt="Image" /></td>
<td><img src="image6.png" alt="Image" /></td>
</tr>
<tr>
<td>26.</td>
<td><img src="image7.png" alt="Image" /></td>
<td><img src="image8.png" alt="Image" /></td>
</tr>
<tr>
<td>27.</td>
<td><img src="image9.png" alt="Image" /></td>
<td><img src="image10.png" alt="Image" /></td>
</tr>
<tr>
<td>28.</td>
<td><img src="image11.png" alt="Image" /></td>
<td><img src="image12.png" alt="Image" /></td>
</tr>
</tbody>
</table>

___

**Anmerkungen/Besonderheiten:**
**APPENDIX C: EXAMPLE OF AN AMBIGUOUS BEHAVIOR TASK**

<table>
<thead>
<tr>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Schau mal, dieser Lupi (zeigen) erzählt etwas. Und das Kind hört ihm zu und weint ein bisschen. Was ist denn da nur los? (Pause) Hat der Lupi etwas böses gesagt, damit das Kind weint? Oder sagt der Lupi dem Kind etwas liebes, damit es nicht mehr weint?</td>
</tr>
</tbody>
</table>