Unpacking the boxes we put people in - On the symmetry, contextual malleability, and maintenance of social categorization

Inauguraldissertation

zur Erlangung des Akademischen Grades

eines Dr. phil.,

vorgelegt dem Fachbereich 02 - Sozialwissenschaften, Medien und Sport

der Johannes-Gutenberg-Universität

Mainz

von

Felicitas Flade

aus Heidelberg

Mainz

2020

Tag des Prüfungskolloquiums: 14. Juli 2020

Erklärung

Chapter 2 beruht auf folgendem zur Publikation eingereichten Manuskript:

Flade, F., Imhoff, R. (2020). *Is social categorization a/symmetrical? Asymmetrical intergroup categorization and individuation become symmetrical by facing category members repeatedly*. Manuscript submitted for publication.

Section 4.1 beruht auf folgendem publizierten Artikel:

Flade, F., Klar, Y., & Imhoff, R. (2019). Unite against: A common threat invokes spontaneous decategorization between social categories. *Journal of Experimental Social Psychology*, *85*, 103890. https://doi.org/10.1016/j.jesp.2019.103890

Felicitas Flade, Mainz, 14. April 2020

Abstract

To categorize (others) is inherently human. Even so, we do not fully understand it yet. Social categorization enables us to structure and understand our social world and helps us save "brainpower". To this aim, social categorization capitalizes on our intuitive grasp of similarity and perceptual flexibility to magnify similarities and differences in our social environment that seem relevant to us. Then again, perceiving others as members of mutually exclusive groups this way often leads to stereotyping, prejudice, and discrimination, widely undesired practices. In the present work, I aim to contribute to our understanding of the "inner workings" of social categorization as a cognitive tool. Do we persistently categorize people of other groups than our own more than our peers? Can our cognitive system take on a life on its own and enter a "vicious circle" of mutually reinforcing categorizations and stereotypes? Can a common enemy weaken our perception of such basic group divisions? These topics are studied empirically and integrated into a discussion about the conceptualization of social categorization in relation to processes of similarity perception and self-identification with a social group.

Keywords: Social categorization, categorization strength, stereotyping, other-race-effect, common enemy, self-fulfilling prophecy

Deutsche Kurzzusammenfassung

(Andere) zu kategorisieren ist zutiefst menschlich. Trotzdem verstehen wir es noch nicht vollständig. Soziale Kategorisierung ermöglicht uns, unsere soziale Welt zu strukturieren und zu verstehen, und dabei "Denkaufwand" zu sparen. Dafür nutzt soziale Kategorisierung unseren intuitiven Begriff von Ähnlichkeit und unsere flexible Wahrnehmung, um Ähnlichkeiten und Unterschiede zu verstärken, die für uns wichtig sein könnten. Andererseits führt das Wahrnehmen von Anderen als Teil einer umgrenzten Gruppe oft zu unerwünschten Nebeneffekten wie Stereotypisierung, Vorurteilen und Diskriminierung. In der vorliegenden Arbeit möchte ich einen Beitrag zum Verständnis von sozialer Kategorisierung und ihrer Funktion als Denkwerkzeug leisten. Kategorisieren wir Menschen, zu deren Gruppe wir nicht gehören, wirklich immer mehr als wir "unsere eigenen Leute" kategorisieren? Kann es passieren, dass unser Denkapparat sich verselbstständigt und sich in ihm Kategorisierung und Stereotypisierung gegenseitig verstärken? Kann ein gemeinsamer Feind diese grundlegenden Grenzen zwischen Gruppen abschwächen? Diese Fragen untersuche ich empirisch und integriere die Untersuchungsergebnisse in eine Diskussion über das wissenschaftliche Konzept von sozialer Kategorisierung im Spannungsfeld zwischen wahrgenommener Ähnlichkeit und Selbstidentifikation mit einer sozialen Gruppe.

Schlagwörter: Soziale Kategorisierung, Kategorisierungsstärke, Stereotypisierung, Other-Race-Effect, Gemeinsamer Feind, Selbsterfüllende Prophezeiung

Contents

Chapter 1 – Introduction

> "Where does the hill start, and where does the valley end? Nowhere! They are one. It is your mind which says, 'This is the valley, and this is the hill.'" - Osho

The real world is dimensional, but our understanding of it is categorical. This is a recurring theme in intellectual discourse: pantha rhei ("everything flows", Heraklit) and natura non facit saltus ("nature makes no jumps", Carl von Linné) convey similar meanings (arguably coined by more reputable thinkers than the one cited above). It is also perhaps one of the few principles of the conditio humana that most academic traditions can agree on (Hirschauer, 2014). Categorization can be seen as precondition of such basic human habits as naming ("all objects with these properties are henceforth called chairs"), and thus of language itself (Malt, Sloman, & Gennari, 2003). More importantly, categorization is likely to be a precondition of all intentional thought in the sense of making meaning of, understanding, expecting events and predicting reactions from our environment. In short, "Orderly living depends upon it." (Allport, 1954, p. 20). A particularly notorious subtype of categorization is social categorization – the sorting of humans into groups (i.e. "putting people into boxes" or "pigeonholing" them (English), "sticking people into drawers" (German) or "putting a label on them" (French)). Generally, social categorization is classified as a special case of (object) categorization. It is assumed to be similarly indispensable to human cognition (Oakes, 2008) and also considered to be based on the perceived relative similarity-dissimilarity between the individuals affected (Bhatti & Kimmich, 2015; Bruner, Goodnow, & Austin, 1956; Hugenberg & Sacco, 2008; Tajfel & Wilkes, 1963). That humans are "social animals" (Aronson, 2002) gives weight to social categorization over and above (object) categorization. Social categorization can have downstream consequences well beyond a better understanding

of the world. Firstly, social categories lead to the formation of psychological stereotypes about others (e.g. “men are always in control of their emotions”), that can contradict their self-image and thus intrude on their right of self-definition (e.g. “I am an emotional man”). Social categories also precede social groups that lend themselves to self-identification (e.g. “I am a woman”; Simon, Hastedt, & Aufderheide, 1997). Self-identification turns social categories into ingroups (“us (women)”) and outgroups (“them (men)”). This has been shown in many studies using the minimal group paradigm (Tajfel, Billig, Bundy, & Flament, 1971). In it, people who are arbitrarily assigned to one of two meaningless groups start treating outgroup members worse than ingroup members, even at the expense of own gains (Otten, 2016; Tajfel et al., 1971). Thus, intergroup settings can easily lead to prejudice, in that people depreciate outgroup members (Bernstein, Young, & Hugenberg, 2007), as well as discrimination, in that individuals are treated worse because they are perceived as members of a certain group (Hewstone, Rubin, & Willis, 2002; Tajfel, 1970). Not least because of these downstream consequences, it is vital to understand social categorization better. Therefore, in the present research, we study how social categorization adapts to different contexts in order to shed light on some of its underlying characteristics. Specifically, we study how social categorization adjusts its *gestalt* to limited intergroup exemplar (= individual category member) encounters, suggest a new mechanism for social category reinforcement and demonstrate that categorization can be reduced by a common enemy. The findings are related to a central tenet of social categorization in the field of social psychology: Its foundation in perceived similarity-dissimilarity between exemplars of different social categories.

1.1 Defining social categorization

For a cognitive tool whose very purpose is definition, social categorization is surprisingly hard to define. In social psychology, definitions of social categorization often use descriptions

of functions and outcomes, so you find some of them in Section 1.2 - Explaining social categorization. While these may be useful contributions to *explaining* social categorization, they only provide us with a vague sense of what categorization *is*. To give us a shared representation of social categorization, I will thus provide my own working definition here. It is partly based on Bruner's (1957, p. 125) definition: "So long as an operation assigns an input to a subset, it is an act of categorization."

I define social categorization as cognitive schema that assigns the same content-free sorting attribute to all individuals within a perceived population subset. Metacontrast (i.e. perceiving similarities between individuals within a population subset as greater than similarities between individuals in the subset and the rest of the population) is a necessary, but not a sufficient condition for social categorization. The content-free sorting attribute, i.e. social category, can be imagined as a blank "tag" (Taylor & Fiske, 1978) or earmark. Everyone is perceived to either have it or not have it (which is the definition of "categorical"). If only one individual had it, it would not add information over and above this individual's unique set of features, while, if everyone had the same "tag", it also would not add information over and above features shared by everyone – in short, in these two cases, categorization is useless. But if only some have it (and, maybe, others share another "tag"), you could fill it in with content, like traits ("nice") or visual features ("small ears"). The attributes thereby assigned to a category are *stereotypes*. Stereotypes themselves do not inherently contain a positive or negative evaluation (i.e. are directly associated with *prejudice*), although many stereotypes are closely linked to an evaluation (Park & Judd, 2005). Stereotypes can, but do not have to map on real-world attributes that vary with categories. "Birds fly" is a stereotype that maps closely on real-world attributes, as most birds fly, and most non-birds do not. As there might be remote reasons for the stereotype that "sharks are killers", there are probably many deadlier fish in the sea. In non-social

categorization, stereotypes that map real-world attribute distributions might be superior to "non-mapping (e.g., invented) stereotypes" in predicting events and actions. The social world, however, is not simply another landscape of "natural" attribute distributions. It is at least as much culturally formed and socially constructed (also including e.g. different cultures in economy and science), so that there can be a fundamental disconnect between the "natural" attribute distributions (e.g. inherent dispositions of people) and the attribute distributions observable in our social environment. When people attempt to make sense of these observable attribute distributions, stereotypes that seem independent of "natural" attribute distributions ("Women do not have business acumen") can become manifested in the structure of the real world. Society rewards behavior consistent and punishes behavior inconsistent with these stereotypes, e.g. rewarding (or not punishing) women who downplay their business acumen by not voicing their good ideas. Thus, the expectation to not hear good business ideas being proposed by women may be(come) confirmed. This can effectively make such stereotypes equally predictive and thus equally functional to attribute-mapping stereotypes in the social domain. While categorization based on an "attribute-mapping stereotype" is already problematic as it biases the prediction of individually different attribute-values towards a categorical prototype, an "invented stereotype" is additionally problematic for suggesting expectations of category exemplars to be someone they are not even approximately "meant to be".

Although every metaphor has its flaws, imagining a social category as an inherently blank "tag"-like attribute makes many of its suggested and studied features and downstream consequences easily conceivable. For example, a tag can easily become a label, when the category is given a name. Also, the blank tag itself, just as anything that is put on it (labels, stereotypes, evaluations) can be a very powerful perceptual equalizer for exemplars of the same category and differentiator between exemplars of different categories. Also, when time

or cognitive resources are limited, one can save energy by taking the "shortcut" of studying the concise tag profile instead of looking at the whole complex person attached to it. Moreover, this way, a social category becomes an (arguably somewhat elaborated) attribute among other attributes of a person. A folk wisdom which has found its way into research on outgroup homogeneity is that when we perceive others in terms of a category, we see them *less as an individual*. But just as we can perceive and process multiple attributes of a person at the same time, it may well be possible to see a person *both* in terms of a category and as an individual at the same time (see Chapter 2). It is conceivable that categories are reinforced when people feel confirmed by the "tag" when they compare it not to the person it is attached to, but their mental image of her or him (see Chapter 3). And it might well be that tags become less important in the face of a common enemy (see Chapter 4).

1.2 Explaining social categorization

Social categorization has been defined as "means of systematizing and ordering the social environment particularly with regard to its role as a guide for action, and as a reflection of social values", or "system of orientation which creates and defines the individual's own place in society", (Tajfel, 1972, p. 293), "cognitive function, which allows for a simplification of perception" (Molenberghs & Morrison, 2014), or "the process of understanding what some thing is by knowing what other things it is equivalent to and what other things it is different from" (McGarty, 1999; Oakes, 2008). As can be seen in these characterizations, social categorization is widely considered to serve two main functions for our cognitive system: saving resources (Fiske & Neuberg, 1990; Molenberghs & Morrison, 2014; Sherman, Macrae, & Bodenhausen, 2011), and structuring our perceived world, so that we can make sense of it (Oakes, 2008; Sherman et al., 2011). As these two needs can therefore be reasons, causes, processes, and aims for social categorization, they are considered *principles* of social

categorization (Rosch, 1978). Two additional shared phenomenological characteristics emerge. Firstly, social categorization is understood as a cognitive process (e.g., Allport, 1954; McGarty, 1999; Molenberghs & Morrison, 2014) Secondly, it is connected to perceiving similarity between some individuals, and dissimilarity between others (Brewer, 1988; Bruner et al., 1956; McGarty, 1999; Oakes, Haslam, & Turner, 1994). While we can legitimately use the attribute "cognitive process" in a definition of social categorization, this is not the case for (increased) perception of similarity-dissimilarity. Not only does it contain a testable proposition about social categorization that should therefore not be written into a definition ("definitional statements are neither true nor false", Markovsky, 2018, p. 50), the concept of similarity itself is ultimately incompatible with a categorical "all-or-nothing" representation (e.g. similarity as "both-and" (sowohl-als-auch), Bhatti & Kimmich, 2015, p. 235). As close(st?) antecedent and / or outcome of social categorization, however, studying perceived similarity-dissimilarity can provide valuable insight into the phenomenon of social categorization (Leeuw, Andrews, Livingston, & Chin, 2016). Moreover, while similarity-dissimilarity might not be a legitimate element in a definition of social categorization in a narrower sense, it nevertheless has been treated that way and thus a consideration of this matter may advance the theoretical discussion on social categorization in the field.

To illustrate the concept of similarity-dissimilarity and its connection to social categorization, imagine twenty same-sized pebbles, ten red, ten blue. Each red pebble is very similar to each other red pebble, but also very dissimilar to each blue pebble. The more the pebbles are all colored in the same shades of red and blue respectively, the clearer it is that there are two "groups" of pebbles: red and blue. This is called the metacontrast ratio. The metacontrast ratio is the mean of all within-category similarities divided by the mean of all between-category similarities. Thus, as within-category similarities increase and/or between-category similarities decrease, the metacontrast ratio increases. The metacontrast ratio has

been used to describe different instances in the categorization process: as antecedent and as consequence. A high metacontrast between stimuli in the information ecology has been considered an antecedent to social categorization, in that it could make a potential category dimension salient. This would then trigger perceptual accentuation by superimposing a categorical structure onto the natural distribution. This view is supported by the prototype model (Rosch, 1978). It states that real features and characteristics are not distributed evenly, but can be correlated and accumulate at certain "points of density" (Rosch, 1978). For example, wings (currently) co-occur more often with feathers than with fur, in what we call "birds", and tall plants usually have wooden stems, while small ones rarely do. There may be functional reasons for such co-occurrences, but this is ultimately irrelevant for the process of categorization itself. Categorization is likely based on "mere co-occurrence" of attributes. However, forming categories along such functional divides may make us aware of them and their potential uses to us. This illustrates that in object categorization, forming categories itself may maximize the outcomes at the cost of barely any side-effects. The same may be true for social categorization, although categorizing other humans may also lead to the substantial side-effects mentioned earlier.

In a reversal of this causal chain from ecological metacontrast to categorization, social categorization is also often characterized as a cognitive amplifier of perceived metacontrast (*accentuation*, Tajfel & Wilkes, 1963), in that social categorization makes perceivers represent a metacontrast that is more pronounced than in the information ecology (Leonardelli & Toh, 2015; Rosch, 1978; Turner, Hogg, Oakes, Reicher, & Wetherell, 1987). For example, in the "Who said what?"-Paradigm (S. E. Taylor, Fiske, Etcoff, & Ruderman, 1978), a measure of social categorization, social categorization is conceptualized as metacontrast ratio in which within-category similarities are greater than between-category similarities.

The notion that metacontrast may be the "key" to social categorization is not limited to social psychology and the cognitive domain. In sociology, it has been suggested that a metacontrast can also be fabricated to reinforce social categories within societal institutions: gender-segregated bathrooms and school uniforms that differ by gender maximize the spatial and visual metacontrast between genders (Hirschauer, 2014). Likewise, similarity as a scientific concept is neither "new" nor restricted to sciences that practice quantitative empiricism (Bhatti & Kimmich, 2015). Cultural studies define similarity as "transformation distance between representations: entities which are perceived to be similar have representations which are readily transformed into one another" (U. Hahn, Chater, & Richardson, 2003, p.1). Bhatti and Kimmich (2015) describe similarity as qualitative proximity, next to spatial, temporal and quantitative proximity. This qualitative and highly flexible "substance" seems to make it difficult for philosophy and cultural science to get a hold of the concept. Similarity itself is described as very intuitive (we just *know* when two things are similar and when they are not). However, its criteria seem so context-dependent (two things can be similar in one context and dissimilar in the next) that similarity evades their conceptual grasp, and the concept remains "vague" (Bhatti & Kimmich, 2015).

Social psychology might not have had this problem, as we can use simplification in study designs to determine a context and choose the dimensions on which similarity judgments can take place (e.g. shape and size of geometrical objects). Alternatively, when social psychologists are interested in ecologically more valid dynamics of similarity judgments, they let the qualitative comparison process happen in the minds of participants and only retrieve the resulting quantitative similarity judgment ("How similar are these two?"). Still, this issue may have contributed to a theoretical controversy on the nature of social categorization within impression formation research in social psychology, but might also help to resolve it. The debate seems to revolve around the relationship between individual exemplar attributes and

attributes associated directly with the category – and the consequences of these views for evaluating social categorization and its functionality. The first perspective is an early social cognitive one (the "piecemeal, elemental, algebraic approach", Fiske, Lin, & Neuberg, 2018). It views attribute features as fixed and summable, so that in theory, a category impression could be computed directly from salient exemplar attribute content (Anderson, 1981). For example, if several individual exemplars were young, short and liked to play, they would be grouped into the category "children". Also, they assumed that attribute meaning would not change by context or reference frame (individual vs. category), so for example, "young" would be always primarily associated with "innocent", so any individual child and also the category "children" would also be associated with "innocent". That individual attributes could predict category attributes in this way was contested by a more "Gestalt, holistic, figural approach" (Fiske et al., 2018) stating that attributes on any level could change their perceived meaning depending on context and prior belief. For example, "young" could mean "did not see 9/11 happen live" when this is the topic of a conversation, or "innocent" when the perceiver has this pre-held stereotype. In agreement with the first view, even if individual attributes can change their meaning, they can still inform category attributes. The "self-categorization theory perspective" (Oakes, 2008) goes even a step further. It claims that there is a fundamental disconnect between individual attributes and category content. Among children in a classroom (categorized as "class"), "young" can mean "lacking knowledge". And while students in the classroom may "lack knowledge", as a "class" they may know more than the class next door and are thus "well educated" on the category level. As individual and category attributes can therefore take on diametrically opposing meanings, this view disconnects the individual and category level content entirely, as if categorical and individual impression formation take place in entirely different dimensions. This fuels their claim that categorization neither simplifies, nor distorts or biases our perception, but only makes

meaning over and above information provided in the ecology (Oakes, 2008). It has been argued that the "piecemeal" social cognitive approach to social categorization and the "holistic" approach assumed by Gestalt theory stand irreconcilable, as attribute meanings either change with context or not (leading to the suggestion of the Continuum model of impression formation, Fiske & Neuberg, 1990). Furthermore, the "self-categorization theory perspective" rejects either perspective for claiming that social categorization produces bias, making a consensus difficult (Oakes, 2008). This debate is of course much more complex than that, so it cannot be dissected in detail here. I believe, however, that cultural studies' "vague" conception of similarity and their experience in dealing with qualitative concepts comes in handy in finding a common ground anyway. Firstly, assigning a "categorical label" to the relation between different attributes might be ill-advised. Does individual attribute content determine stereotype content or not? Are attribute meanings fixed or can they be changed until they essentially become another attribute? Probably neither but a little of both, and likely to different degrees. This variance may have interpersonal, intrapersonal and content-related components that invite empirical investigation - and not a quest for definitional authority. Secondly, the contributors to this debate may have confused qualitative similarity with quantitative similarity. The attributes "lacking knowledge" and "well-educated" are quantitatively highly dissimilar. On a scale from "no knowledge" to "very knowledgeable", they are located at opposite ends. From the perspective of qualitative similarity, however, they are very similar. They share many features ("about knowledge", "quantify a performance", "about an ability to acquire cognitive representations"), but only require a single transformation (Bhatti & Kimmich, 2015) to bridge the gap between the two: "learn more". That we perceive these two attributes as very dissimilar can be attributed to the ease with which we find a difference, and paradoxically, this becomes easier the more similar they objectively are (*structural alignment*, Gentner & Markman, 1994). The structural

alignment view states that processes of comparison rely on finding a basis of comparison first. When comparing chairs and tables, we do that based on them both being furniture. Then, we proceed to find differences ("sit on" vs. "sit at", "carry people" vs. "carry food"). Structural alignment posits that there are two kinds of differences: alignable and unalignable. Alignable differences refer to attributes that are shared across objects that are compared, such as in the example above. Chairs and tables both have to do with sitting, but in different ways. Unalignable differences do not refer to such shared attributes, e.g. a chair and the sky: a chair is wooden, the sky is not. Indeed, unlike alignable differences, unalignable differences seem to be mainly restricted to function, parts, category, and material (Gentner & Markman, 1994). Gentner and Markman (1994) found that participants faced with similar stimulus pairs could list *both* more similarities and alignable differences than participants faced with dissimilar stimulus pairs. This might help us understand metacontrast and accentuation in relation to social categorization. Social categorization might not just be the differentiation between "incomparable" social groups. We may first have to notice that they are both human in order to make differences. This could also be one way in which social categorization makes meaning and saves resources at the same time. Indeed, alignable differences hold much more value for meaning making, and they require highly similar entities. At the same time, perceiving many differences between similar entities seems to come naturally and may thus require a minimal amount of cognitive resources.

1.3 Characterizing social categorization: The research program

So far, I laid out the most central functions of social categorization and proposed a definition. To describe a phenomenon, however, we need not only know what social categorization *is* and what it is *for*. We also need to know what it is *like*, its nature or character. How does it function? Does it adapt its functionality to context changes? What

maintains, what threatens its stability? Recent research suggests that social categorization can occur automatically and spontaneously (Weisman, Johnson, & Shutts, 2015). As such, it seems to neither require cognitive resources (Sherman et al., 2011) nor motivation (Brubaker, 2007; Sherman et al., 2011). Moreover, categorizing along learned category dimensions does not require a metacontrast in the information ecology (*intercategory fit*, Wegener & Klauer, 2005), although intercategory fit does enhance categorization strength (Dotsch, Wigboldus, & van Knippenberg, 2011). Regarding many other "characteristics" of social categorization, social psychologists seem to hold a range of beliefs that remained largely untested so far. For example, social categorization is considered hardly malleable (with the exception of cross-cutting categories, Klauer, Hölzenbein, Calanchini, & Sherman, 2014), perhaps also due to some scholars including all cognitive representations into their definition of social categorization (*categorization as representing*, Klapper, Dotsch, van Rooij, & Wigboldus, 2017). Also, depending on research tradition, researchers believe that social categorization is asymmetrical (Hugenberg, Young, Bernstein, & Sacco, 2010; Park & Rothbart, 1982) or symmetrical (Klauer et al., 2014; Kurzban, Tooby, & Cosmides, 2001), and that it is independent of self-identification (Tajfel & Wilkes, 1963) or dependent on it (Kawakami, Amodio, & Hugenberg, 2017). Theoretical definitions define a concept, boundary conditions define the underlying phenomenon. If we as a field believe that social categorization as a coherent phenomenon is "real" in that it exists independently of social construction (maybe even in contrast to other fields), we must aim to empirically study the phenomenon and correct our concepts accordingly, and not vice versa. This way, we might isolate the phenomenon from its manifestations in specific category content such as gender or race, so it can be studied to tell us something about the workings of the human mind – and allow us to generalize across generations of human minds. In the present work, I hope to take some tentative steps into that direction. We investigate instantiations of symmetry change in social

categorization (Chapter 2), its interaction with stereotyping to perpetuate categorical representation (Chapter 3) and categorization reduction by a common threat (Chapter 4).

Regarding the empirical investigation of social categorization, we concentrated on social categorization strength in the present work. Categorization strength, however, is just one of the dimensions of social categorization. Other researchers have also investigated the shrinking and expanding of a category to include more or less exemplars with marginal categorical fit (*category inclusivity*, Dovidio, Gaertner, Hodson, & Houlette, 2004), vertical changes in categorization level within a category hierarchy (Gaertner et al., 2000), and exemplar identification (selection of one over another category for categorization) as determinants of social categorization (Ito & Urland, 2003). The factors studied here with respect to their influence on social categorization strength may well have effects on these related outcomes, too (e.g. Chapter 4, recategorization under common threat), and the definition suggested above is designed to also include these dynamics (one could ask, e.g., How relevant is the "tag" to forming an impression about an individual? (When) is a "tag" assigned to this person? Which "tag" is assigned?). Of these outcomes, however, categorization strength seems to be the least studied and one of the most relevant to understand the content-free cognitive schema that is social categorization.

Here, it may also be worth noting that from the viewpoint of humanities, social categorization can manifest in many more ways than usually imagined by social psychologists. For example, a third "diverse" gender is introduced to government forms, when urban sub-cultures develop their own dressing codes, when dialects of the same language become separate languages, when someone verbally differentiates people by assigning them to differently labelled groups – all of this can be called social categorization. Acts like labelling or making categories visible by dress codes may well be antecedents and consequences of social categorization in the social cognitive terminology adopted here (Bigler

& Liben, 2007). These practices may also use tools like accentuation or structural alignment that directly appeal to cognitive social categorization, so it might be compelling to subsume these phenomena under the same term. But for now, and especially for the present thesis, we might be well-advised to delineate our own (social psychological) social categorization and categorize it, along with related phenomena such as labelling or institutionalizing discrimination along category borders, into the superordinate category of "human differentiation" (Hirschauer, 2014).

To study a phenomenon such as social categorization strength, that often occurs automatically and spontaneously (Weisman et al., 2015) and thus likely often outside our awareness, a specialized paradigm is required. As the "Who Said What?"- Paradigm is the primarily used measure in the present work across all empirical chapters, it is introduced here.

1.4 The "Who Said What?"- Paradigm as measure of categorization strength

The "Who Said What?"- Paradigm (WSW, S. E. Taylor et al., 1978) is the current state-of-the-art measure of automatic social categorization strength. It is based on a memory task and thus consists of two phases: In the encoding phase ("discussion phase"), 8 "speakers", 4 from each category (e.g. black and white US Americans), are presented sequentially paired with statements. Each speaker is presented 6 times, and every trial features a new statement, resulting in 48 subsequent presentations of binary speaker-statement pairs. In the recall phase ("assignment phase"), all statements (plus as many new statements) are presented again, and participants have to choose for each statement, which of the eight speakers "said" it – or whether it was not presented previously. The main logic behind this paradigm is that participants, confronted with a sentence they cannot reallocate to the correct speaker, may use a speaker category attribute as proxy to increase their chance at guessing the correct speaker (S. E. Taylor et al., 1978). For example, they might not remember that Jack said it, but that the

speaker's skin color was light. If this is the case, the participant should randomly choose a speaker from the same category for their answer – resulting in more within-category errors. This is traditionally assessed by the error-difference measure that compares the sums of within- and between-category errors (S. E. Taylor et al., 1978). A higher within- than between-category error rate would be attributed to the application of social categories in the memory task. As participants are not explicitly informed about "categorization" being the variable of interest, the paradigm is considered to be a relatively unobtrusive measure of social categorization (Klauer & Wegener, 1998). Also, no stereotypes are provided for the categories, so attributes distinguishing between categories can be selected, weighted and applied naturally and unprompted. This makes the captured process highly ecologically valid. Despite these desirable facets of the WSW paradigm, there are some issues with the interpretation of the classical error score.

Klauer and Wegener (1998) pointed out several sources of noise in the statistical operationalization. They argue that an answer in the WSW task can stem from different cognitive pathways, which is ignored in the classic analysis. For example, if statements are not remembered at all, participants are forced to guess. Thus, in that situation, the probability to choose an answer option is equal for each of the answer options. In the classic analysis, the resulting pattern of error frequencies (same amount of within- and between-category errors) enters the analysis as evidence for "no categorization", while the pattern was caused at an earlier stage of the cognitive process. This could lead to an underestimation of actual categorization strength. To be able to estimate this old/new memory, Klauer and Wegener (1998) introduced (new) distractor statements in the recall task and added a "new" answer option. Similarly, giving a correct answer for a previously encountered statement can not only be the result of explicit speaker memory, but also correct guessing on the basis of category memory for that speaker, or even random guessing. Thus, Klauer and Wegener (1998)

proposed a multinomial processing tree (MPT) model of social categorization which accounts for these shortcomings. The model can tease apart contributions of individual cognitive processes, substantially decreasing random error in the processes of interest and thus increasing the power to detect the effect. The model parameters reveal imperfect statement memory (D), the share of variance occupied by exclusive person memory (C), the amount of applied category memory (d) and possible biases in category- or old-new (expectancy-based) guessing (a, b). Thus, the WSW task can distinguish between and measure person memory (individuation) and category memory (categorization strength) largely independently. This allows us to predict and test effects on categorization strength exclusively, e.g. controlling for individual person memory.

All MPT analyses were performed by means of Bayesian hierarchical latent-trait MPT modelling (Klauer, 2010), which allows for taking interindividual variability into account, thereby improving model fit and allowing for correlating parameter estimates. It is based on two steps of data augmentation and uses Bayesian methods with a weakly informative hyperprior distribution and a Gibbs sampler.

Chapter 2 – On the a/symmetry of social categorization

Abstract

Is social categorization asymmetrical, as in the other-race effect (ORE), or symmetrical, as frequently found by means of the "Who said what?"-Paradigm (WSW)? We traced social categorization within the different methodological constraints of these two paradigms. In a reanalysis of previous studies from our lab, we established the symmetry of intergroup categorization and individuation in WSW data patterns (Study 2.1, $N = 1212$), and showed that this symmetry is not a methodological artefact of the WSW design (Study 2.2, $N = 81$). In Study 2.3 & 2.4 ($N = 99 / 88$), we aimed to reduce the Black-White ORE in the classical face perception task by decreasing the number of exemplars per category and presenting exemplar stimuli repeatedly. In Study 2.5 ($N = 112$), we adapted the WSW paradigm to accommodate the ORE, in order to study its sub-processes ingroup and outgroup categorization and individuation. Intergroup categorization and individuation asymmetries both become symmetrical under repeated category member exposure. This finding contributes substantially to a comprehensive conceptualization of social categorization.

> "Wer als Spezialist für die Kategorisierung von Menschen anhand äußerer Kennzeichen auftritt, muss sich selbst als außerhalb dieser Kategorien präsentieren, als ungefärbt, unmarkiert. Kurz, Hautfarben, die über sie Auskunft geben, haben immer nur die anderen."
>
> ("Anyone who acts as a specialist for the categorization of people based on external characteristics must present himself as being outside of these categories, as uncolored, unmarked. In short, it is always the others who have skin colors that tell you about them.")
>
> – Valentin Groebner, Austrian historian (2003)

When Sam enters the classroom to see her new class for the first time after the summer break, she notices small groups of students: the rich kids, the cheerleaders, the geeks, and the outsiders, and trouble abounds. This is how a typical high school film might start off. Much research suggests that the real-life scene might be bleaker - the teacher might simply be more prone to categorize the other-race kids than the own-race kids, and not see any of the other colorful "groups" at all. If Sam is white, she might see the black kids mostly as "black" at first, while already starting to notice individual differences between the white kids. In the present research we aim at contrasting these two perspectives and their corresponding research traditions by systematically dissecting the procedural details that are responsible for results in line with one or the other.

The "pop culture" notion (sorting all individuals in mutually exclusive "boxes") and the scientific finding (categorizing and grouping the other race, individualizing the own race) represent two fundamentally different perspectives on categorization: Whereas the former depicts categorization as a symmetrical process of individuals either ending up in one box (category) or the other, the latter construes categorization as inherently asymmetrical. In current theorizing on intergroup categorization, while the asymmetrical approach is more recognized and elaborated, the symmetrical one is implied in certain research traditions, too. Both perspectives implicitly claim that the symmetry assumption underlying their notion of social categorization is the default one. This might be because neither perspective feels vulnerable – both have very good arguments. On the one hand, we confuse people from (racial) outgroups more strongly than ingroup members (Chance & Goldstein, 1996). To European White people, Black people seem more alike, while Black people may have a hard time telling Asians apart (a phenomenon labelled the Other-Race-Effect; for a review, see Meissner & Brigham, 2001). On the other hand, social categorization reliably seems to be equally strong for in- and outgroup when measured by the "Who said what?" – Paradigm (S.

E. Taylor et al., 1978), irrespective of the category dimension examined. The present research takes a look at these two empirical approaches to basic categorization of individuals into groups and seeks to elucidate if and why categorization is typically asymmetrical in the former but not in the latter approach.

2.1 Asymmetrical categorization and the ORE face perception task

The most prominent asymmetrical phenomenon rooted in social categorization is the robust "principle of outgroup homogeneity" (Park & Rothbart, 1982, p. 1051). It is the "apparent tendency for within-group accentuation of similarity to apply to outgroups rather more than it does to ingroups" (Oakes, 2008). Outgroup homogeneity denotes the tendency to see outgroup members as more mutually alike than ingroup members. A plethora of tasks were designed to measure this phenomenon, only a few of which produced stable asymmetry effects under meta-analytic scrutiny (Boldry, Gaertner, & Quinn, 2007). Of the eleven investigated measures, only two showed the patterns they predicted for outgroup homogeneity effects. We chose the most prominent one, the face perception task, as outgroup homogeneity measure for the present research.

More precisely, the face perception task does not comprehensively capture outgroup homogeneity, but merely its purely visual equivalent – the Other-Race-Effect (ORE). The ORE – recognizing ingroup member faces better than outgroup member faces – is one of the most eminent (Feingold, 1914) and stable (Meissner & Brigham, 2001) effects studied in social psychology. In the classical face perception task used to study the ORE, participants are presented with previously seen and unseen black and white portraits. White US Americans falsely mark more new black portraits as seen than new white portraits. In signal detection theory terminology (Green & Swets, 1966), participants thus show *lower sensitivity* for outgroup faces. To explain this effect, a common notion in ORE theorizing is that perceivers

tend to think categorically about outgroup exemplars, while processing ingroup members in a more individuated manner (Hugenberg et al., 2010; Young, Hugenberg, Bernstein, & Sacco, 2012). This notion is directly opposed to findings obtained with another measure of social categorization: The "Who said what?"-Paradigm (S. E. Taylor et al., 1978).

2.2 Symmetrical categorization and the WSW paradigm

The "Who said what?"-Paradigm (WSW, S. E. Taylor et al., 1978) is a widely recognized unobtrusive measure of social categorization strength. Participants are presented with statements by black and white US American speakers (or speakers from any other dual categories). Subsequently, when asked to reassign statements to speakers, participants commit more within-category-errors than between-category-errors. The magnitude of this error-difference is considered a measure of spontaneous social categorization strength. Most published WSW studies make use of intergroup settings: Most participants could easily identify with one of the presented speaker categories. For example, female and male participants take part in a female-male WSWs, Black and White US participants take part in WSWs featuring the same speaker categories (Flade, Klar, & Imhoff, 2019; Kurzban et al., 2001).

Findings by means of the WSW paradigm include the malleability of social categorization by a competing category (Klauer et al., 2014) or a common enemy (Flade et al., 2019), the application of the stereotype dimensions agency and progressiveness to categorize occupations (Imhoff, Koch, & Flade, 2018), and the application of the trustworthiness dimension to categorize faces (Klapper, Dotsch, van Rooij, & Wigboldus, 2016, but see Degner, Imhoff, & Dunham, 2020). The Multinomial Processing Tree Model of social categorization (Klauer & Wegener, 1998) can disentangle WSW data into parameters representing probabilities of cognitive processes by analyzing frequencies in all obtained

distinct response categories, including the frequencies of reassigning the correct speakers depending on categories. These parameter estimates include the independent probabilities of using category memory and individual person memory. The majority of data patterns obtained by means of the WSW paradigm show equally strong perception of individuals in terms of their category membership between categories (Kurzban et al., 2001; S. E. Taylor et al., 1978), even leading to default equality constraints between parameters d_a and d_b (Flade et al., 2019; Klapper et al., 2016; Klauer et al., 2014, for an exception, see Imhoff et al., 2018). This should be the case particularly when categories appear as binary contrasting pairs on (opposite ends of) the same category dimension. In everyday language, this is often the case, as we talk about each other, and also in contemporary research: Male and female, black and white, old and young (Leonardelli & Toh, 2015). Thus, here, the tendency to perceive an exemplar in terms of their category membership is the same regardless of ingroup / outgroup or White vs. Black category. Although not thoroughly discussed in the literature to this point, in its current scope, this notion of intergroup categorization as symmetrical by default may contradict theoretical deliberations on outgroup homogeneity and empirical findings of asymmetrical categorization.

2.3 Comparing WSW and ORE paradigms

In intergroup settings, the WSW study setting and design are very similar to those of the ORE face perception task. Procedurally, participants bring their individuation motivation and experience in telling apart individuals' faces into a study setting featuring portrait stimuli from two distinct categories: one that can be considered their (most often racial) ingroup (e.g. White US Americans, White Germans), and the largest intranational outgroup (e.g. Black US Americans, Turks). In both tasks, equal numbers of exemplars from both categories are

presented to the participants in a concealed memory task with automated stimulus presentation in the encoding phase and a forced choice task in the recall phase.

Theoretically, both approaches discussed here have defined categorical processing as "distinguishing between groups without necessarily distinguishing between their members" (Klapper et al., 2017, p. 4), also referred to as the "grouping" definition of social categories (Klapper et al., 2017). Both mentioned approaches (and we in this paper) treat the phenomenon of categorization as one where individuals make sense of their social surroundings by lumping persons into neat categories, like "men", "Blacks" or "elderly" (for a discussion of other approaches to categorization than this "grouping" aspect see Klapper et al., 2017). It is determined by measuring the extent of perceptual "lumping" of exemplars of the same category and "splitting" of exemplars of different categories (cf. accentuation principle, Tajfel & Wilkes, 1963). Both paradigms used in this work therefore operationalize social categorization by means of confusing category members with each other by either ascribing statements of one category member to another (WSW) or mixing up seen and unseen category members (ORE). Specifically, exemplar confusion *within* one category is compared to between-category confusion (in the WSW paradigm) or confusion within the other category (in the ORE face perception task).

In both the WSW and the ORE, the dependent variable can be explicated as "perceiving individuals in terms of their category membership rather than individually" (Klapper et al., 2017). Yet, the ORE is the *stronger* perception of *outgroup* exemplars in terms of their category membership relative to ingroup exemplars, while WSW data commonly does not show such an asymmetry. Both the signal detection analysis of the face perception task and the MPT analysis of the WSW task principally allow for both symmetrical and asymmetrical categorization. The similarities and easily assessable differences between WSW and ORE

theorizing and operationalization provide an intriguing setting for theoretical and empirical advances in studying social categorization.

The present research

To explore the symmetry of intergroup categorization and the role of categorization in the ORE more specifically, we employed a method-driven approach and made use of the paradigms' design constraints. We conducted a reanalysis of previous studies (Study 2.1), two studies varying the classical ORE face perception task (Studies 2.3 and 2.4), and two studies using the "Who said what?" – Paradigm (Studies 2.2 and 2.5). In Study 2.1, we empirically explored whether there are asymmetries in intergroup categorization or individuation in previous studies from our lab featuring the standard WSW paradigm. In Study 2.2, we investigated whether asymmetrical categorization could appear symmetrical in the WSW paradigm due to design constraints, as non-members of one category in the paradigm (e.g. the outgroup category) could be encoded as a "non-categorical" category. While we critically examined the symmetry of WSW categorization in this study, in Studies 2.3 and 2.4 we put the asymmetry in the ORE to the test by varying the number of category exemplars and stimulus presentation repetitions. Specifically, in Study 2.3 we aimed to decrease the ORE by decreasing the number of exemplars per category. In Study 2.4, we aimed to decrease the ORE by presenting the exemplar stimuli repeatedly. In Study 5, we adapted the WSW paradigm based on the previous studies to accommodate the ORE, in order to separately measure its sub-processes ingroup and outgroup categorization, and ingroup and outgroup individuation. This research line aims to provide new insights into the symmetry and nature of spontaneous social categorization, and its contribution to the other-race-effect. We report all studies conducted in this research line and therein all measures, manipulations, and exclusions. Final sample size was always determined before data collection. All materials,

data and supplemental analyses are available on our OSF project site https://osf.io/avunp/?view_only=c618c7a5e67f42508f51729f0c96195a.

2.4 Study 2.1

Previous findings suggest that categorization is usually symmetrical in the WSW paradigm, but this impression has not been tested empirically yet. In this study, we empirically investigated categorization symmetry in the WSW paradigm by means of a reanalysis. As there are similar theoretical considerations for individuation, and categorization and individuation are closely interrelated in the ORE face perception measure (Hugenberg et al., 2010), we investigated individuation symmetry in the same manner. Thus, we reanalyzed all previous WSW studies collected in our lab that featured interracial settings and portraits as speaker stimuli.

Method

Sample of Studies. We included all (published and unpublished) studies from our lab (all carried out 2015 – 2019) featuring White and Black US American, as well as White German participants and matching categories (Black and White) in the speaker stimuli ($N = 9$), featuring a total of $N = 1212$ participants. This subset of studies was chosen for conceptual closeness to the most robust other-race-effect, the one between White and Black category exemplars.

Inclusion criteria. While self-identified ethnicity was collected in US samples, study participants from ethnic minorities (especially Black participants) at German universities are so rare that ethnicity data was not collected and all participants in these studies were treated as White. As all of the studies featured two between-participant conditions unrelated to the question at hand and most of them included White US American and Black US American

participants, we split all datasets by Black/White participant ethnicity and condition, resulting in $k = 34$ independent datasets to be included in the analysis.

Results

As estimate of categorization, we used Cohen's d_z as ES metric derived from *t*-tests between stimulus categories on the difference score between within- and corrected between-category error frequencies. For individuation, we used Cohen's d_z as ES metric derived from *t*-tests between stimulus categories on hit frequencies. As for categorization, despite considerable heterogeneity, Q (33) = 111.40, $p < .001$; $I^2 = 70.4\%$, the random-effects model did not indicate that categorization was stronger for the participant-specific outgroup relative to the ingroup ($d = 0.00$, 95% - CI [-0.17–0.17], $z = -0.01$, $p < .99$; Fig. 2.1). The results for individuation were comparably heterogeneous, Q (33) = 98.36, $p < .001$; $I^2 = 67.16\%$, but provided no indication that individuation was stronger for the participant-specific outgroup relative to the ingroup ($d = 0.13$, 95% - CI [-0.02–0.29], $z = 1.66$, $p < .10$). Thus, neither categorization nor individuation seem to be asymmetrical in race intergroup settings as measured by the WSW task and analyzed by the error-difference measure.

2.4.3 Discussion

Across our previous studies, we did not find asymmetrical categorization or individuation in intergroup categorization when measured by means of the WSW paradigm featuring Black and White categories. Although we only used studies from our own lab, we confirmed data patterns previously published by others (Klapper et al., 2016; Klauer et al., 2014). Notably, the findings from this study are limited to the WSW paradigm and its constraints. Some of these constraints could also be at work in the real world and thus inform our understanding of context-dependent dynamics of social categorization symmetry. Next, we investigated one constraint that could force categorization into symmetry in the WSW paradigm.

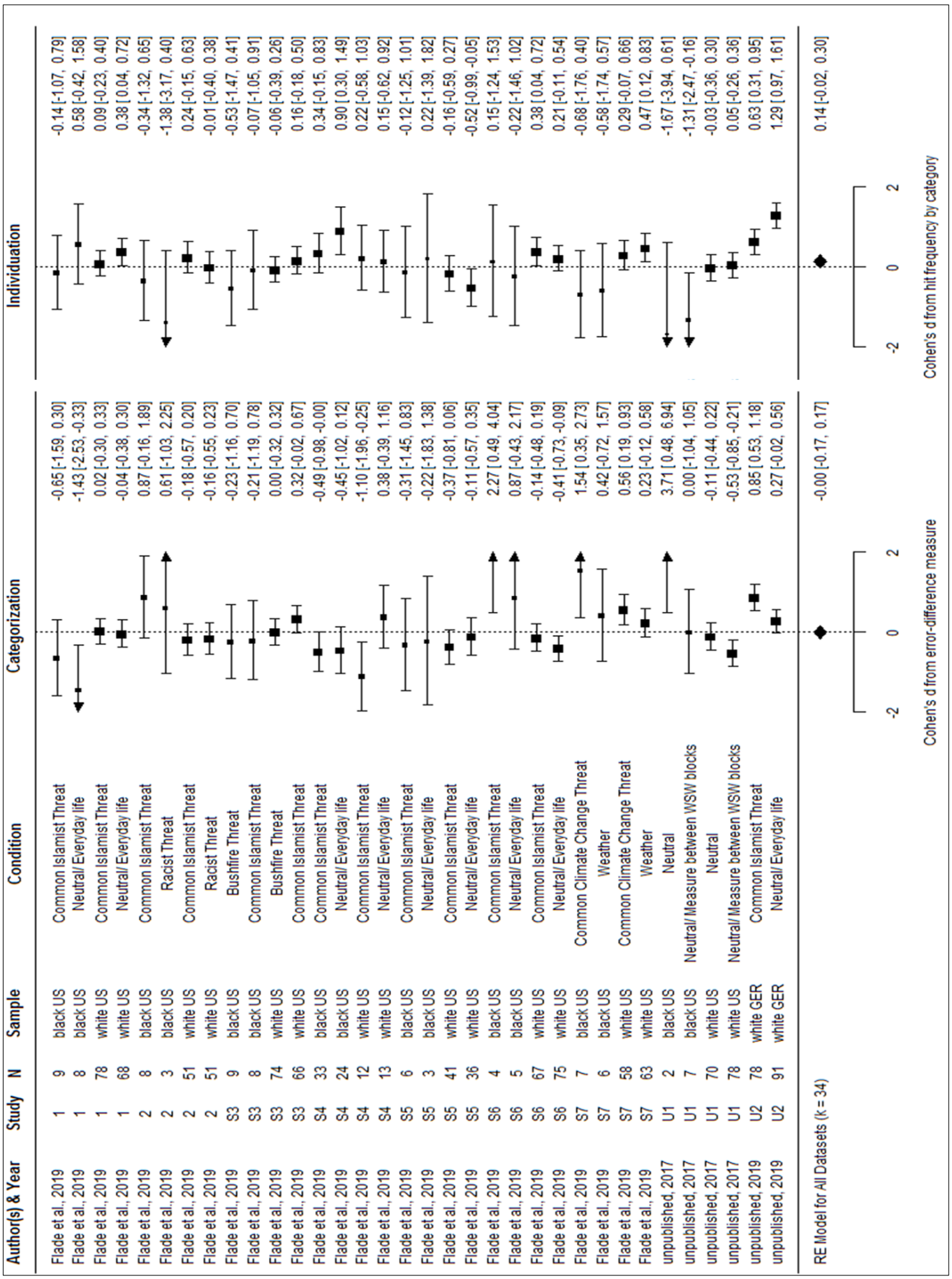

Figure 2.1. Forest plot for reanalysis in Study 2.1 with random - effects model.

2.5 Study 2.2

Study 2.1 did not provide any evidence for asymmetrical categorization. This could mean that ingroup and outgroup are always lumped together to the same extent in making sense of social information ecologies. There is, however, an alternative explanation. It is conceivable that participants perceive only one group as an actual category (and thus categorize asymmetrically), but recode the task in a sense whereby all other speakers become defined by not belonging to this one focal homogenous category. Thus, as the number of categories is limited to two in a standard WSW paradigm, the actual lack of categorization of a group of non-categorical exemplars may be concealed by the artificial category of "non-categorical exemplars" (e.g. "non-black" as opposed to a black category). Therefore, only encoding one of the categories and all other exemplars as "non-categorical" could lead to the same result as encoding two categories independent from one another. The confusion between these non-categorical exemplars would then be measured as categorization, although they are not actually perceived as a category, but participants merely remembered that the speaker was "not Black". Does this also drive the symmetrical categorization measured in the WSW paradigm? We conducted this study to exclude this construct-unrelated alternative explanation. The study was preregistered under http://aspredicted.org/blind.php?x=6eu25r.

Method

Participants. One hundred US-Americans took part in the study on Amazon Mechanical Turk in exchange for $2.50. An automatic filter only allowed them to participate if they had not participated in any previous WSW study conducted by the authors' lab. As preregistered, if participants indicated at the end of the study that they either saw their data not fit for analysis ($n = 5$) or that they had taken notes during the experiment ($n = 14$), their data were not analyzed. Thus, the data of 81 participants (42 men, 39 women, $M_{age} = 36.45$, $SD_{age} = 11.18$, 62 White, 2 Hispanic/Latino, 10 Black/African American, 1 Native American/

American Indian, 3 Asian/ Pacific Islander, 2 White-Hispanic/Latino, 1 White-Asian/Pacific Islander) were included in the analysis. Power analysis is not yet available for the hierarchical Bayesian implementation of MPT models in the R package 'TreeBUGS' (Heck, Arnold, & Arnold, 2018), therefore, we determined a-priori sample sizes by compromising between the current standard in the social categorization literature using MPT analysis and new standards in the field of social psychology. Thus, we report here post-hoc sensitivity analyses for achieved power in the error-difference measure. The present study had 80% power to detect an effect size of $d_z = .32$ on the classical error-difference measure.

Stimuli/ Manipulation. Four portraits of black Americans were randomly drawn from a pool of 8, and one portrait of a White, Asian, Arab and Latino American each were randomly drawn from pools of 3 for each participant anew. All portraits were chosen from the Chicago Face Database (Ma, Correll, & Wittenbrink, 2015). The portraits displayed faces with a neutral expression which scored highest on the respective races' pre-rating in the CFD coding manual ("Other race" for the Arab category). The statement set was designed to feature neutral and race-irrelevant content like "I like reading books" or "I have a daughter". See online supplementary material for complete statement set and CFD portrait names.

***Procedure and Hypothesis*.** After accepting the HIT, participants accessed the study via a link to the SoSci Survey platform (Leiner, 2014), where they gave informed consent and performed the WSW task. They were instructed that they were about to see several "young people meeting for the first time and engaging in a dialogue". Then, the participants were presented with successive paired presentations consisting of a speaker and a statement each. Statements were randomly assigned to speakers irrespective of category membership. The speaker was presented first and for 9 s, and the statement was displayed after a 1.5 s delay, so both stimuli were then simultaneously displayed for 7.5 s. There was no inter-trial break before the next stimulus pair was presented. After observing all 48 pairings, participants

moved on to the surprise recall task. In the surprise recall task all statements from the presentation phase (48) and distractor set (48, in total 96 statements) were shown in random order, and participants were asked "Who said that?" each time. They responded by ticking one of nine answer options, namely the eight portraits and the option "None. This statement is new." For exploratory purposes, participants then indicated the similarity of each binary pair of portraits. Then, participants were debriefed and asked to indicate their age, gender and ethnicity as well as questions about their perceived data quality and whether they took notes during the study.

If the symmetry apparent in Study 2.1 was caused by the fact that participants recode the available information into one salient category (Blacks) and a non-Black category, we would expect symmetric categorization even under these circumstances. In other words, people should then as willingly misattribute a White speaker's statement to an Asian or a Latino speaker as they do for Black speakers' statement to other Black speakers. If, on the contrary, we do observe asymmetric categorization under these circumstances where only one category is a category proper, whereas the other is an eclectic mix of identities, this would speak against the notion that people just recode the task. Moreover, any use of the non-Black category (even in the case of overall asymmetrical categorization) would tell us that the WSW could be biased towards symmetrical categorization.

Results and Discussion

While the error-difference-measure descriptively indicated higher categorization strength for the black US American category (M = 1.43, SD = 3.94) than the portraits of varying race (M = 1.00, SD = 2.39), the difference was not significant, $t(80) = -.75$, $p = .45$, $d_z = -.07$ CI_d [-.38, .24]. As the error-difference-measure may over- or underestimate the real effect, however, we preregistered the MPT analysis as relevant analysis, given appropriate model fit.

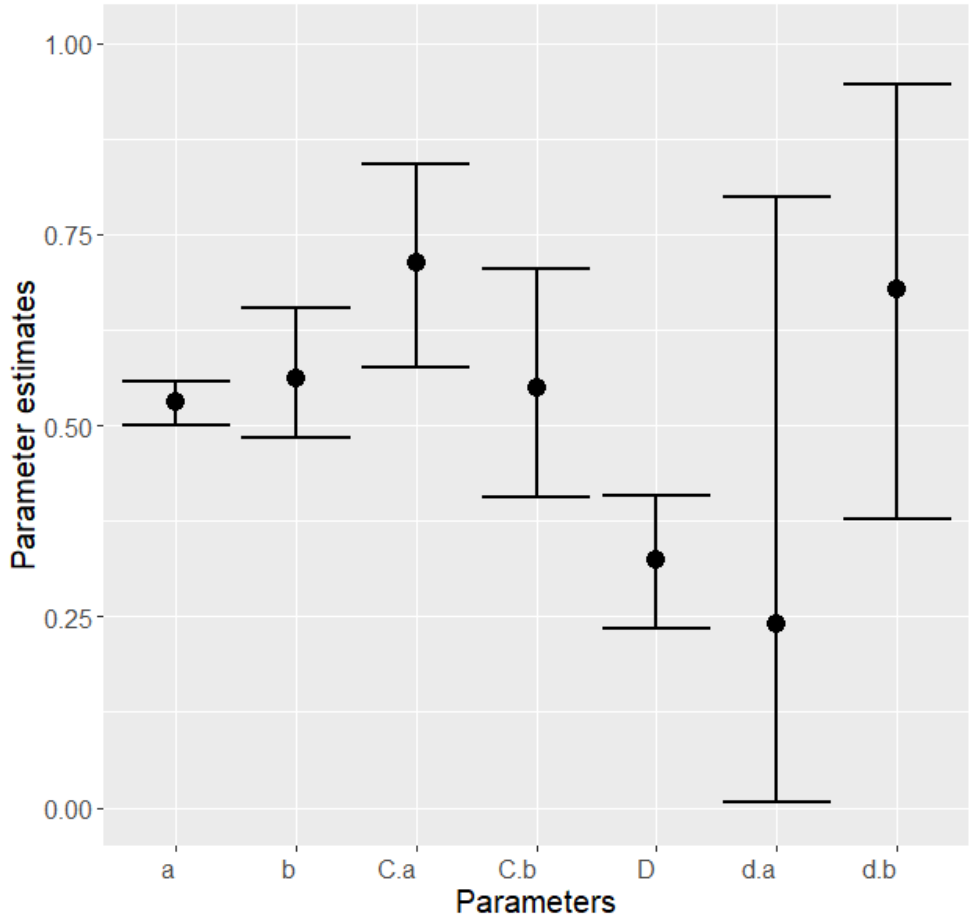

Figure 2.2. Mean parameter estimates and 95% CIs (credibility intervals) in Study 2.2. Subscript a denotes parameter values for the non-categorical speakers, subscript b denotes parameter values for Black category exemplars.

Model fit was appropriate when categorization parameters d_a and d_b were let free to vary ($T_1^{observed} = 0.144$, $T_1^{predicted} = 0.087$, $p = .13$, $T_2^{observed} = 13.05$, $T_2^{predicted} = 15.87$, $p = .65$; results displayed in Fig. 2.2), but broke down when restricting both parameters to be equal ($T_1^{observed} = 0.179$, $T_1^{predicted} = 0.087$, $p = .049$, $T_2^{observed} = 13.84$, $T_2^{predicted} = 15.87$, $p = .61$). Therefore, members of the black US American category were categorized significantly stronger than the portraits of varying race. Restricting the categorization parameter for the portraits of varying race (d_a) to be zero maintained an appropriate model fit ($T_1^{observed} = 0.144$, $T_1^{predicted} = 0.089$, $p = .15$, $T_2^{observed} = 13.44$, $T_2^{predicted} = 15.62$, $p = .61$). Thus, it can be concluded that the portraits of varying race were not significantly categorized in the WSW

paradigm. This indicates that the WSW paradigm does not artificially trigger and measure the categorization of "non-category-exemplars", rejecting the hypothesis that the symmetrical categorization observable in the WSW paradigm is (even partly) caused by the design. Having validated the measure in this regard, we tested next whether the ORE would hold when making the face perception task more similar to the WSW.

2.6 Study 2.3

Is categorization asymmetrical in the ORE face perception task because a large number of category exemplars is presented in each category, possibly binding more cognitive resources? In the classical face perception task, the number of presented exemplar portraits per category usually ranges from 12 to 50. Contrary to the relatively low number of exemplars in the WSW (typically 4 per category), this may force participants to save resources in the face perception task and prioritize some faces (e.g., easily distinguishable ingroup faces) over others. If participants have a finite number of individual memory slots and fill these more readily with exemplars from their own category, they have less capacity left for other-race faces and might therefore exhibit an ORE. Specifically, they will engage in decreased individual encoding of unfamiliar outgroup faces. In order to test this hypothesis, we conducted an ORE face perception task with only 4 instead of several dozen portraits per category (as is typical for the WSW). Under these circumstances, it seems unlikely that asymmetric performance would be the result of processing resources being limited by the number of exemplars (Kareev, 2000).

Method

Participants. Ninety-nine German students' full datasets were obtained in the lab (10 men, 89 women, $M_{age} = 23.69$, $SD_{age} = 5.18$). No datasets were excluded. This study had 80% power to detect an effect size of $d_z = .28$ for the difference in sensitivity between categories.

Stimuli/ Manipulation. Eight portraits of black and eight portraits of white US Americans were randomly drawn from pools of 30 for each participant anew. Four of each selected set were presented to the participants, four served as distractors. All portraits were chosen from the Chicago Face Database (Ma et al., 2015). The portraits displayed faces with a neutral expression which scored highest on the respective races' pre-rating in the CFD coding manual.

Procedure. Participants took part in this study as part of a larger set of unrelated studies in the lab at the University of Mainz. They were presented with 4 black and 4 white randomly selected portraits for 2 s each in random sequence. Then, they were presented with those portraits again, and as many distractor portraits, and had to indicate for each portrait whether they had seen it previously. After indicating their age and gender, they were debriefed.

Results and Discussion

The face perception task was analyzed by means of Signal Detection Theory's sensitivity index d′: The higher the hit rate (old faces recognized as old) and the lower the false alarm rate (new faces considered as old), the better the detection of the signal. Typical ORE are manifested in asymmetric d′s: the sensitivity index is significantly higher for ingroup than for outgroup faces (Malpass & Kravitz, 1969). Our reduced design also yielded an ORE (white ingroup: $M_{d'} = 1.41$, $SD_{d'} = .33$; black outgroup: $M_{d'} = 1.16$, $SD_{d'} = .43$; $t(98) = 5.28$, $p < .001$, $d_z = .62$, 95% - CI_d [.33, .90] ; ROC-curve see Fig. 2.3). Thus, the difference in categorization symmetry between ORE and WSW is not fully due to number of presented category exemplars. The second major difference between the two paradigms is stimulus repetition. Therefore, we assimilated the face perception task in this respect to test this aspect in Study 2.4.

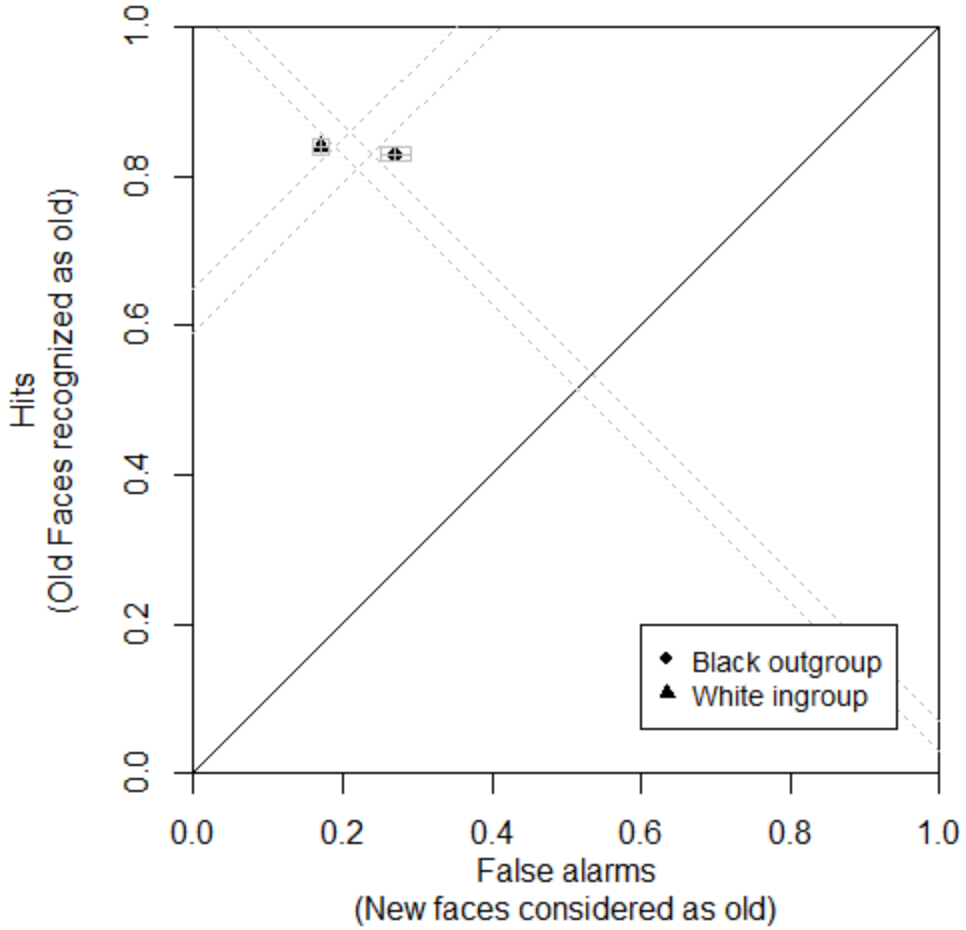

Figure 2.3. Receiver operating characteristic (Mean d′ and 95% CIs) for Study 2.3. Sensitivity for Black and White faces is plotted as a function of Hits (frequency of old faces recognized as old) and False alarms (frequency of new faces falsely considered as old). The drawn through diagonal denotes performance that is not different from chance: There are as many Hits as False alarms. The area under the diagonal (indicating negative performance) is thus typically not needed. The opposite diagonal denotes sensitivity: The more Hits and the fewer False alarms, the better recognition performance in this task. As can be seen from the point estimates (auxiliary lines along the confidence intervals for better readability), the memory for ingroup faces is more sensitive (falling diagonal) and participants had a more conservative response criterion (rising diagonal) towards ingroup faces.

2.7 Study 2.4

Is categorization asymmetrical in the ORE because each category exemplar is presented only once? A single presentation might limit the available cognitive resources during the presentation phase, as participants' attention is diverted by the unusual features of outgroup faces from building recognition memory for individual exemplars. These "unusual features"

are likely to be common features within a single outgroup category and to align with the category dimension, making categorization of these faces more likely (MacLin & Malpass, 2001). The lack of repeated presentations could preclude participants from making up for this initially biased recognition memory. Thus, lack of perceptual expertise in participants regarding outgroup faces in general could lead to weaker individual outgroup face memory in the face perception task. We tested whether the ORE would recede when altering the Study 2.3 design to include exemplar stimulus repetitions. This study was preregistered under http://aspredicted.org/blind.php?x=22gx4m.

Method

Participants. One-hundred-and-one US American participants took part in this study on MTurk. As preregistered, participants were excluded when indicating that they did not see their data fit for analysis (n = 1), or that they took notes during the study (n = 3), or that they already participated in an ORE task (n = 6; multiple: n = 3). Eighty-eight datasets were included in the analysis (54 men, 33 women, 1 other, M_{age} = 33.95, SD_{age} = 9.02, 67 White, 3 Hispanic/Latino, 11 Black/African American, 5 Asian/ Pacific Islander, 1 White/Black, 1 White/Asian). Two participants did not produce hits in in one or both categories, so we could not compute both d′ estimates for them, resulting in them missing from the respective analyses. This study had 80% power to detect an effect size of d_z = .30 for the difference in sensitivity between categories.

Stimuli/ Manipulation. Same as Study 2.3.

Procedure. Participants took part in this study on MTurk. They were presented with 4 black and 4 white randomly selected portraits for 2 s each in random sequence, each portrait was shown 6 times. This is the typical number of repetitions in the WSW paradigm. Then, they were presented with those portraits again, and as many distractor portraits, and had to

indicate for each portrait whether they had seen in previously. After indicating their age and gender, they were debriefed. They received $0.70 for their participation.

Results and Discussion

In this design, we did not find an ORE anymore (white exemplars: $M_{d'}$ = 1.30, $SD_{d'}$ = .56; black exemplars: $M_{d'}$ = 1.21, $SD_{d'}$ = .55; $t(85) = 1.96$, $p = .053$, $d_z = .22$, 95% - CI_d [-.08, .52]; ROC-curve see Fig. 2.4). The pattern persisted when analyzing the data separately for participants self-identifying as either only White (white ingroup: $M_{d'}$ = 1.28, $SD_{d'}$ = .61; black outgroup: $M_{d'}$ = 1.21, $SD_{d'}$ = .57; $t(64) = 1.41$, $p = .16$, $d_z = -.06$, 95% - CI_d [-.41, .28]) or only Black (white outgroup: $M_{d'}$ = 1.47, $SD_{d'}$ = .28; black ingroup: $M_{d'}$ = 1.47, $SD_{d'}$ = .18; $t(10) < .01$, $p > .99$, $d_z < .001$, 95% - CI_d [-.80, .80]). Thus, the ORE was not reliably detected anymore if stimuli were repeated.

First of all, these results are mute as to what causes ORE as they are compatible with both a motivation and a perceptual expertise perspective (Hugenberg, Miller, & Claypool, 2007). It might be that during a single presentation, the outgroup category features are so salient that they override any attempt at the processing of individual features. This is in line with previous research suggesting that repeated exposure attenuates the other-race-effect (Markant & Scott, 2017) and that other-race faces are processed more slowly (Markant & Scott, 2017; Natu, Raboy, & O'Toole, 2011). Multiple presentations may enable people to get used to the category features, making these features less salient and letting the perceptual system focus on individual features. For the present research, Studies 2.3 and 2.4 could establish the boundary conditions within which the ORE asymmetry should also be detectable by means of the WSW paradigm.

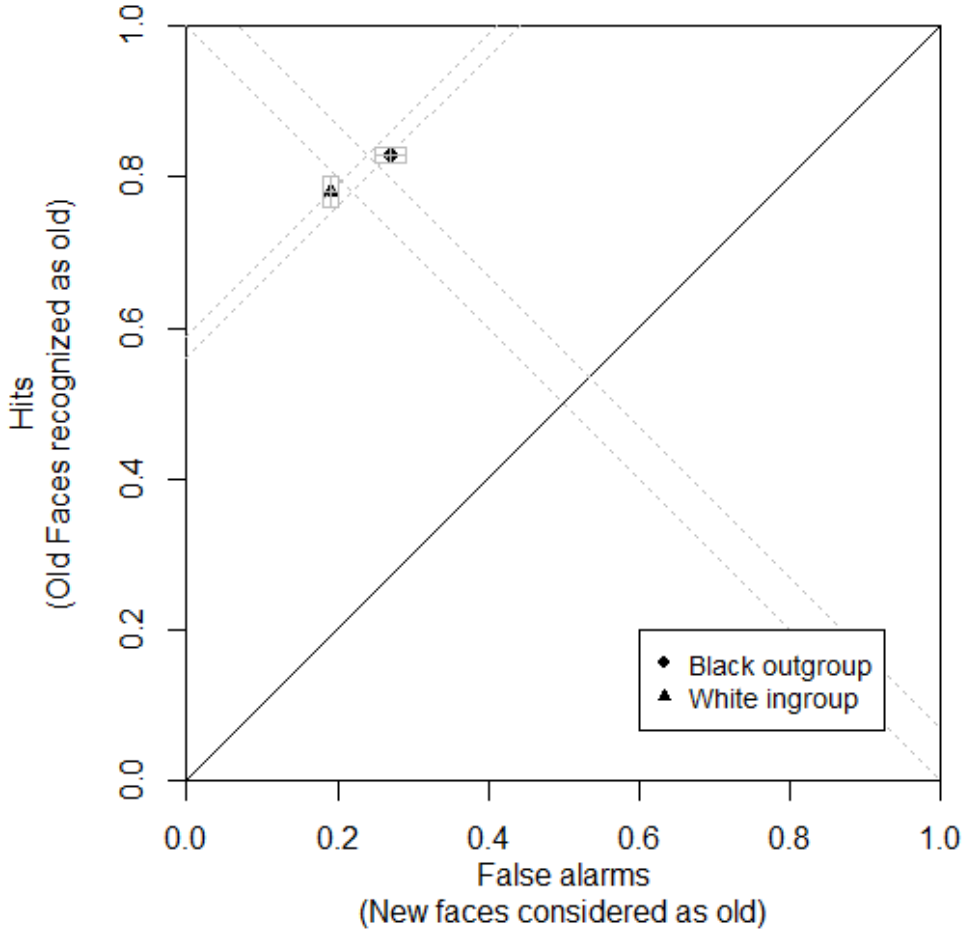

Figure 2.4. Receiver operating characteristic (Mean d′ and 95% CIs) for Study 2.4. Sensitivity for Black and White faces is plotted as a function of Hits (frequency of old faces recognized as old) and False alarms (frequency of new faces falsely considered as old). The drawn through diagonal denotes performance that is not different from chance: There are as many Hits as False alarms. The area under the diagonal (indicating negative performance) is thus typically not needed. The opposite diagonal denotes sensitivity: The more Hits and the fewer False alarms, the better recognition performance in this task. As can be seen from the point estimates (auxiliary lines along the confidence intervals for better readability), the memory for ingroup and outgroup faces is equally sensitive (falling diagonal) but participants had a more conservative response criterion (rising diagonal) towards ingroup faces.

2.8 Study 2.5

In Study 2.4, we established that the ORE can be significantly reduced by repeatedly presenting ingroup and outgroup faces. This might also be the reason why there is no reliable asymmetry on either individual person memory or category memory in our previous WSW studies along the race dimension. In Study 2.5, we aimed to validate the conclusions from our

previous studies and seize the opportunity to study categorization and individuation symmetry separately, as well as the relation between social categorization and individuation.

Theoretical deliberations on social categorization rarely go without discussing its relation to individuation. Indeed, we need to understand the relation between categorization and individuation in order to accurately interpret patterns of a/symmetry by either. To this point, all three kinds of relationship between categorization and individuation could be imagined: A negative dependency, in that higher categorization comes along with lower individuation, a positive dependency, in that higher categorization comes along with higher individuation, or an independent relationship. Categorization asymmetry across categories could therefore determine a mirroring individuation asymmetry and/or vice versa. If they are independent, one of the two could be symmetrical while the other is asymmetrical.

Turning to prominent modes of person perception, the Dual Process Model of Impression Formation (Brewer, 1988) describes two subtypes of categorization and two subtypes of individuation. Here, automatic "primitive categorization" by race, gender or age within a multidimensional space is followed by an evaluation of relevance and self-involvement, leading to feature-based category subtyping. Thus, one could use the model threefold to explain categorization asymmetry. An asymmetry could already happen at the initial categorization level ("Identification"), in that a category is detected for outgroup members, while it is not detected for ingroup members. Alternatively, race might be detected (symmetrically) for both ingroup and outgroup during this initial stage, but as self-involvement is present for ingroup members, own-race exemplars are individuated ("personalized"), while self-involvement is absent for outgroup members, so other-race exemplars are categorized or subtyped. Lastly, both ingroup and outgroup could be both "identified" and "subtyped" within their race categories. But as more information is available top-down to be attributed to ingroup exemplars, they are individuated more than outgroup

exemplars. In the first explanation, individuation would not occur at all. In the second, ingroup members would be personalized and outgroup members categorized (and then possibly individualized). This would make categorization and individuation qualitatively distinct processes, that are also interdependent within categories – the more an ingroup member is personalized, the lower her categorization, the more an outgroup member is subtyped, the more she is individualized. In the third explanation, categorization and individuation would also be fully interdependent within the same category. Thus, this model seems to favor interdependency of categorization and individuation within a category.

The second explanation is also in line with Hugenberg et al.'s (2010) Categorization-Individuation Model that seems to suggest that categorization (stronger for outgroup faces) and individuation (stronger for ingroup faces) are two potentially related albeit distinct processes ("the tendency for categorization […] can be overridden via motivation to focus on individuating characteristics", p.1170; "*categorization* and *individuation* […] two qualitatively different ways of attending to and encoding social targets", p. 1170). Similarly, the early version of the Continuum model of impression formation (Fiske & Neuberg, 1990) posits that individuation and categorization lie on the same continuum: increasing individual processing should decrease categorical processing and vice versa – they should be strongly negatively related. This corresponds to the measurement in the ORE task. The single measure d′ is interpreted as *both* a measure of individuation and categorization at the same time, which implies that the ORE necessarily includes both an individuation asymmetry and a categorization asymmetry. On the contrary, the MPT model of social categorization models individuation and categorization parameters separately and thus allows them to be largely mutually independent.

Implementing our previous findings to accommodate an ORE-like asymmetry in the WSW paradigm could help resolve some of these inconsistencies. If the ORE is strongest

when stimuli are only presented once (i.e. “first encounter” in the real world) and does not just affect recognition memory but also source (category) memory, it should also be present in a WSW without stimulus repetition. As individuation and categorization are conflated in the ORE signal-detection measure, the WSW implementation also enables us to investigate whether this asymmetry is due to differing individual person memory and / or category memory as a function of ingroup vs. outgroup category. The present study addressed this by implementing a WSW paradigm with 48 different (non-repeated) speakers rather than the standard six-fold presentation of the same eight speakers. This study was preregistered under https://aspredicted.org/blind.php?x=mr3x2q.

Method

Participants. One hundred and twelve students of the University of Mainz took part in the study at the lab in exchange for €7. No datasets were excluded based on the preregistered exclusion criteria. Thus, the data of 112 participants (39 men, 73 women, M_{age}=21.85, SD_{age}=2.50) were included in the analysis. We determined a-priori sample size as in Study 2.2. Thus, we report here post-hoc sensitivity analyses for achieved power in the difference between error-difference measures by category. This study had 80% power to detect an effect size of d_z = .24 between the error-difference indicators of categorization strength.

Stimuli/ Manipulation. Twenty-four portraits of black Americans and 24 portraits of white Americans were randomly drawn from pools of 30. All portraits were chosen from the Chicago Face Database (Ma et al., 2015). The portraits displayed faces with a neutral expression which scored the highest on the respective races’ pre-rating in the CFD coding manual and lowest on all other races’ pre-ratings. The statement set was designed to feature neutral and race-irrelevant content like “I like cooking” or “My neighbors are quiet”. See online supplementary material for complete statement set and CFD portrait names.

Procedure. After registration, participants accessed the study at the lab computers, where they gave informed consent and performed the WSW task. They were instructed that they were about to see "48 people accompanied by a quote taken from them". Then, the WSW task proceeded as described in Study 2.2. This time, they responded in the assignment phase by ticking one of 49 answer options, namely the 48 portraits and the option "None. This statement is new." Then the participants were debriefed and asked to indicate their age, gender and about their perceived data quality.

Results and Discussion

The error-difference-measure indicated higher categorization strength for the black category (M = 9.37, SD = 3.94) than for the white category (M = 6.94, SD = 4.02), $t(111)$ = 5.20, $p < .001$, d_z = .50, 95% - CI_d [.23, .76]. Hit frequencies indicated higher individuation for the white category (M = 6.09, SD = 3.52) than for the black category (M = 3.59, SD = 2.43), $t(111)$ = -8.86, $p < .001$, d_z = .75, 95% - CI_d [.48, .1.02]. The two measures were negatively correlated ($r(112)$ = -.32, p = .001), seemingly suggesting a partial trade-off between individuation and categorization. As the sum of answers for one category is fixed, however, frequencies in response categories are statistically interdependent. Thus, a high hit rate restricts the possible error difference in that category much more than a low hit rate. Therefore, we corrected for the method-determined variance by dividing the error-difference by the sum of within- and between-category errors. This reduced the correlation to insignificance ($r(112)$ = -.10, p = .30).

As the error-difference-measure may over- or underestimate the real effect, we preregistered the MPT analysis as relevant analysis, given appropriate model fit. To test the hypotheses, d_a and d_b, and C_a and C_b were not restricted to be equal, respectively. The model did not fit the data perfectly ($T_1^{observed}$ = 0.152, $T_1^{predicted}$ = 0.068, p = .04, $T_2^{observed}$ = 13.20, $T_2^{predicted}$ = 8.46, p = .15). Nevertheless, the parameter estimates confirm the results of the first

analysis. Categorization was stronger for the black category (M_{da} = .71, SD_{da} = .04) than for the white category ((M_{db} = .40, SD_{db} = .07); Δ d = .31 with the 95% credibility interval [.15, .50], p_B < .001; see Fig. 2.5). Individuation was stronger for the white category (M_{Cb} = .30, SD_{Cb} = .02) than for the black category ((M_{Ca} = .16, SD_{Ca} = .01); Δ d = .14 with the 95% credibility interval [.10, .18], p_B < .001; see Fig. 2.5).

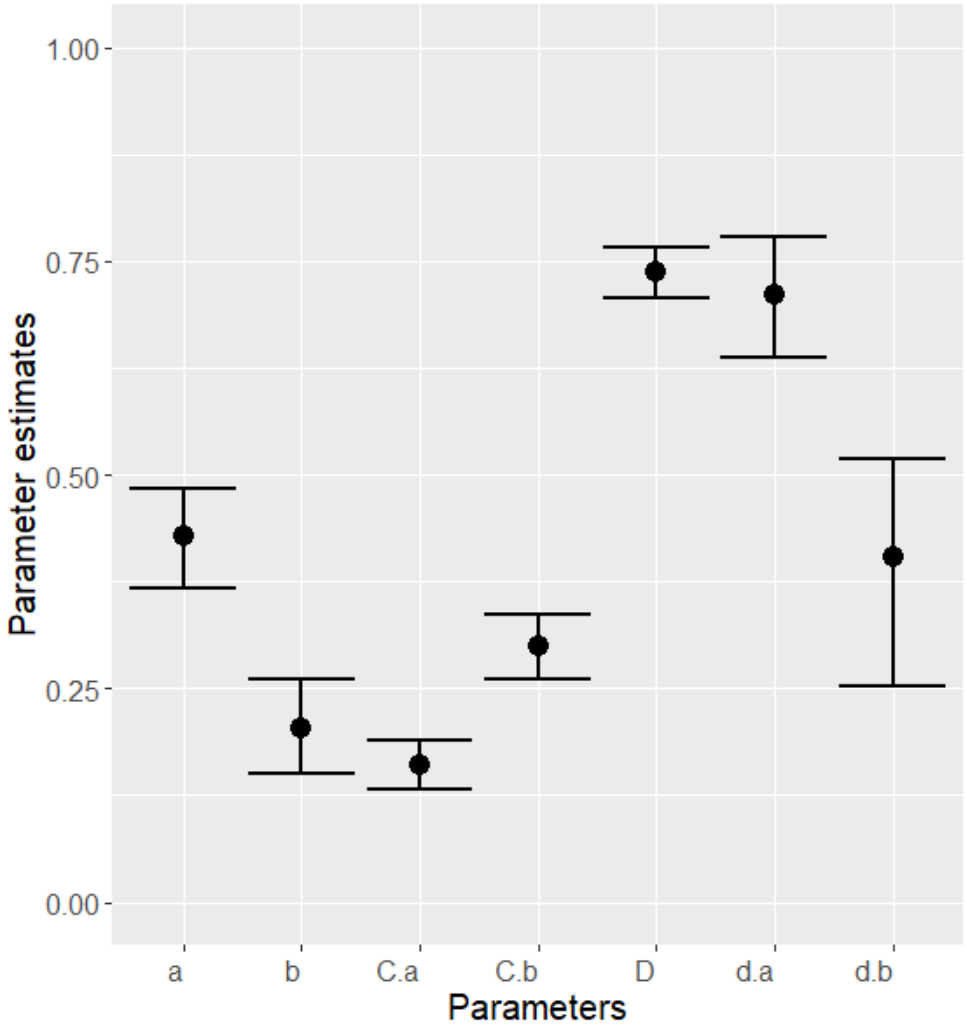

Figure 2.5. Mean parameter estimates and 95% CIs (credibility intervals) in Study 2.5. Subscript a denotes parameter values for the Black category, subscript b denotes parameter values for the White category.

Thus, ORE-like asymmetrical effects replicated within the WSW paradigm. Asymmetries showed on the categorization memory and individual person memory measures / parameters. This suggests a clear asymmetry in intergroup categorization, indicating the need for extending our theoretical understanding of the symmetry dynamics of social categorization over time.

On the other hand, there was significant and considerable categorization of both ingroup and outgroup members. Categorization was significantly higher than individual person memory in any of the categories. Moreover, the correlation between individuation and categorization was insignificant after correcting for method variance. Thus, the results do not show a trade-off between categorization and individuation at the individual level, as implied by the Continuum Model, or a reversal of the dominant process between groups, in that we (predominantly) categorize outgroup members and individuate ingroup members. Both hypotheses could not previously be tested by means of the classical ORE paradigm.

This is not easily alignable with the notion that individuation and categorization are qualitatively different, yet fully interdependent. While the first notion is shared by both WSW and ORE traditions, ORE theories often also assume the second one. That two qualitatively different processes that both "positively" construct meaning could be fully interdependent seems only plausible under strong resource limitations, e.g. finite cognitive and time capacity. To reiterate the tag metaphor for social categorization from the introduction, during a single encounter, there might only be time to glance either at the social category "tag" or the individual attached to it. Yet, it might be that such strong resource limitations are the reason for the asymmetry of early social categorization (and individuation), but that the associated correlation between them was not significant due to large interindividual variability of the resource limitation required.

Interestingly, finding an ORE-like asymmetry in the WSW paradigm also indicates that this asymmetry could not only apply to recognition memory, but also to source memory. This could extend the application of the ORE from eyewitness situations to the possible proliferation of category labels in everyday communication: If it is more likely that we remember individual speakers from our ingroup but more likely that we remember the category membership of an outgroup member, this might be reflected in our own account of events and thus contribute to the everyday salience of outgroup categories.

2.9 General Discussion

Intergroup race categorization and individuation appear asymmetrical when ingroup and outgroup exemplars are encountered initially but become symmetrical under frequent exposure. We established this empirically by showing that the ORE asymmetry diminishes in the face perception task when exemplars are presented repeatedly, while it shows in the WSW paradigm when exemplars are presented only once.

On a methodological level, similar manipulations produced similar data patterns in the face perception task and the WSW. This reinforces measurement validity for both paradigms and indicates that both measure the same construct: social categorization. Naturally, this only applies to social categorization as defined in Chapter 1, explicated by the perceived metacontrast that it produces (confusing exemplars within one or more distinct categories), and measured by these measures' operationalizations. These categories are encoded and represented by within-category attributes that might be used for e.g. task-solving purposes. Social categorization furthermore appears to follow the same rules in recognition memory (face perception task) and source memory (WSW). Notwithstanding its stability across these cognitive domains, social categorization may be expressed differently depending on context and available resources. Our results suggest that asymmetrical categorization (and

individuation) occur under limited exposure frequency. This might prevent observers from getting accustomed to the "alien" - and therefore initially salient - outgroup exemplar features, which might inhibit observers' ability to focus their attention on individuating exemplar features (Levin, 1996; MacLin & Malpass, 2001). Such a process might also explain the lack of evidence for a reliable reduction of the ORE through contact (Hugenberg et al., 2007; Hugenberg et al., 2010). Encountering people who visibly belong to another ethnic group might draw attention to their "otherness" or outgroup features irrespective of liking of, friendship to, or closeness to outgroup members. Regular encounters of the same outgroup member might be necessary, but not sufficient in reducing the ORE for previously unseen outgroup members. To sustainably train our cognition to encode and process individuating features of previously unseen outgroup members, it might be necessary to have regular encounters - well beyond a few hours of laboratory training - with large numbers of new outgroup exemplars, possibly combined with the motivation or need to individuate (such as teachers of mixed-race schools or accountants at public authorities).

While the ORE might play a major role in the specific, but important area of early social categorization of newly or incidentally encountered ethnic outgroup members, social categorization in most other contexts may still be symmetrical. Even in Study 2.5, though clearly asymmetrical, categorization (and individuation) for ingroup and outgroup were of similar magnitudes. This contradicts the notion that ingroups are mainly individuated while outgroups are mainly categorized (Hugenberg et al., 2010; Young et al., 2012). Moreover, while self-identification might indeed alter the symmetry of (early) intergroup categorization, early social categorization without self-involvement is still likely to exist - and to be symmetrical. Our results are not clearly interpretable with respect to the categorization-individuation trade-off. While this trade-off seems plausible in the aggregated data pattern of

Study 2.5, there is no negative correlation between the two, which contradicts the individual trade-off hypothesis.

2.9.1 Limitations & future directions

Due to our method-driven approach to the symmetry of social categorization, our main aim was to "translate" the designs of the ORE face perception task and the WSW "into each other". Thus, our results only speak to the relative symmetry of categorization at zero or five stimulus repetitions. It would be interesting to narrow down these boundary conditions further: Does a single repetition already even out intergroup categorization? Can a single but very lengthy exposure have the same effect, or one in which individual processing is motivated? Similarly, does repeated exposure increase ingroup categorization or decrease outgroup categorization to reach symmetry? As a side note, both paradigms are blind to the attributes or features that individuals use to form categories. Instead, researchers select stimuli in a way that seem to be representing the category of interest (e.g., "Black" and "White" faces). This may introduce ambiguity whether categorization detected in such measures generalizes to ecologically more valid environments, as the stimuli arrangements usually introduce variance on only one, with few exceptions maximally two dimensions (for a similar argument regarding stereotype dimensions see Imhoff & Koch, 2017; Koch, Imhoff, Dotsch, Unkelbach, & Alves, 2016). Thus, a more bottom-up approach to stimulus selection would be needed to properly address the former questions. Owed to the method-driven approach, the conceptualization of social categorization adopted by these measures is quite narrow and technical, and in no way incorporates the manifold theoretical elaborations thought out in social psychology and beyond. Integrating these would require substantial additional work beyond the scope of this paper. The present research nevertheless aims exclusively at informing the theoretical conceptualization and understanding of social categorization. Therefore, the high application value of the ORE remains untapped in the present work. That

social categorization may often be symmetrical does not mean that its downstream consequences, such as prejudice and discrimination, also have to be. They can still be asymmetrical – due to unequal perception and experience of, self-identification with and motivations towards the members of the resulting different categories (cf. Categorization-Individuation Model, Hugenberg et al., 2010).

The successful application of the WSW paradigm to the ORE and its ability to distinguish between ORE individuation and categorization components suggests further avenues of inquiry. For example, existing interventions might have a stronger impact on individuation or categorization selectively and thus be differentially effective depending on context. De-categorization of outgroup exemplars may block the transmission of negative and positive inferences from group stereotypes to new outgroup members and therefore reduce pre-judgmental expectations and discrimination towards them. At the same time, de-categorization might prevent the spreading of positive experiences with individual category exemplars to other category exemplars. On the other hand, de-individuation of ingroup members might also reduce the ORE, but at the cost of side effects like reducing persuasiveness of ingroup members (Wilder, 1990) and decreasing attribution of mind to ingroup members (Deska, 2018).

Is outgroup categorization a figure against the background of "us", or is it the same as ingroup categorization, and our own ingroup membership is immaterial? Maybe neither. Outgroup categorization might be more like an illusory giant – much larger than ingroup categorization at first sight, but as we look closer, just as large as its ingroup counterpart. The same might be true for Sam the teacher. While she may be prone to categorizing only black students at first, as she gets to know her class during term, she will see all students in her classroom equally – both as part of groups and as individuals.

Chapter 3 – Category reinforcement by construal bias

Social categorization as a process as well as many concrete category dimensions are surprisingly persistent determinants of human cognition and societies. While the last chapter was concerned with dynamics of "real-time" category application, this chapter proposes a mechanism whereby social categories may be carried forward inter-individually and across time. This chapter consists of two parts. In the first part, I define and situate the phenomenon of stereotype-consistent interpretation of ambiguous information (i.e. construal bias) theoretically, while the second part contains an empirical investigation of the influence of construal bias on category reinforcement.

3.1 Construal bias: Defining and situating a (not so) novel phenomenon

There has always been a certain fascination with the propensity of human minds to jointly engage in self-perpetuating circles, for the worst (or sometimes the best) of their unwitting owners. Such "vicious circles" were given fittingly dramatic names such as self-fulfilling prophecy or "reign of error", Rosenthal effect, Pygmalion effect, Andorra effect, and maybe also Boudon's (1981) "cumulative processes" as part of the Vicious Circle of Poverty. All these phenomena have in common that a cognition (e.g. prejudice against women in the work context) leads to an act that might contradict one's own explicit values and behavioral ideal (not employing a woman because of her gender). That leads to the target conforming to stereotypes (not being the main breadwinner), which feeds back into the first prejudice and cements the position of this social group in society. The same processes do the same for high-status groups such as financial elites – here, privilege is preserved. These processes are usually promoted by a vacuum in the information ecology: due to the lack of counterevidence,

beliefs or schemata developed in the past (or offered by a third person, as in many related experimental studies) can "fill the gap" and affect thought and behavior (cf. Oeberst & Imhoff, 2020). We situate construal bias in this class of phenomena. We define *construal bias* as the representation of a mental object associated with a target person (observed or reported behavior, traits or events) that is enriched by - and therefore biased towards - a stereotype about the target person. For example, when a Frenchman, a German and an Israeli tell you exactly the same thing – that they really like fresh bread – you might imagine (and "remember" later) that they really like fresh baguette, whole grain rye bread, and pita, respectively. Construal bias occurs via a process we name *stereotype imputation*. Stereotype imputation is substituting a gap in the information on observed or reported behavior, traits, or events with imagined (Slusher & Anderson, 1987) details that are consistent with at least one category attributed to the target person. Therefore, construal bias is constrained by (a) the diagnosticity of the stereotype dimension suggested by the salient ambiguity for a target-relevant category (Skowronski & Carlston, 1989) (e.g., one can quite accurately infer a person's nationality from the bread variety they usually consume) (b) the extent of ambiguity implied by the information ecology (e.g., bread-baking traditions vary widely by country), and (c) the accessibility of category-consistent stereotype content that can replace the ambiguity: if the ambiguity points to "prototypical" stereotypes, it is more likely to instigate construal bias (e.g., bread-baking traditions are a central element in many national prototypes; cf. accessibility x fit formulation of category salience, Oakes, 1987).

Notwithstanding many established neighboring concepts, we argue that construal bias is warranted as an additional distinct concept. Self-fulfilling prophecies are defined over their effect on the observable world, and the affected behavior of the target specifically. For example, students' performances increase based solely on altered teacher belief (Rosenthal & Jacobson, 1968). Conversely, construal bias does not necessarily have any effect on the target,

the effect could remain restricted to the perceiver. Just like construal bias, spontaneous trait inferences are described as “unintended, unconscious, and relatively effortless inferences of traits” (Uleman, Adil Saribay, & Gonzalez, 2008, p. 331). During spontaneous trait inferences, a single concrete observed behavior develops into an impression of a stable trait of the target person. This can be measured by the increasing abstractness of descriptions applied to that person, which signifies that they become more informative about the person, less informative about a specific situation, more enduring, less verifiable and more disputable (Linguistic Category Model; Semin & Fiedler, 1991). Together with a special case, the Linguistic Intergroup Bias (that could be described as difference in spontaneous trait inference based on intergroup bias, Maass, 1999; Sherman, Klein, Laskey, & Wyer, 1998), spontaneous trait inferences are based on external information, even if it might be a minimal amount. That information is ascribed truth value by definition and points towards (or away from) a stereotype (i.e. stereotype-(in)-consistent observed behavior). That runs counter to the information gap required for stereotype imputation that is contained in stereotype-relevant but not -(in-)consistent observed behavior. The same is true for theoretical and empirical arguments made in the area of impression formation in general: the presence and not the absence of external information that can be interpreted in line with a stereotype is considered the starting point of these processes (Kunda & Thagard, 1996). Thus, in the following studies and future studies in this research line, we aim to fill this conceptual gap and establish construal bias as another subtle mechanism that can trigger a similar(ly) vicious circle.

3.2 Making all the difference: Stereotype-consistent interpretation of ambiguous statements reinforces social categorization

Abstract

Information in the real world, and communication particularly, is inherently incomplete and ambiguous. Previous research suggests that ambiguous statements are likely to be interpreted in line with stereotypes about their speakers. We propose that this phenomenon contributes incrementally to perpetuating categorization. We suggest that ambiguous expressions invite perceivers to impute stereotypes into information gaps. This results in an informational enrichment that biases the meaning of the original statement content towards the category prototype, creating what we term *construal bias*. Construal bias would therefore constitute a subtle process of category reinforcement based on a purely cognitive feedback loop. We tested whether construal bias increases social categorization in three preregistered Studies (N = 267). In Studies 4.1 and 4.2 (N = 60 / 100) German and Syrian immigrant speakers were categorized stronger when uttering more construable statements. In Study 4.3, we generalized the effect to the age category dimension (N = 107). The effect is discussed regarding its contribution to understanding the stability of social categories and stereotype maintenance.

> "@ilduce2016: "It is better to live one day as a lion than 100 years as a sheep." - @realDonaldTrump #MakeAmericaGreatAgain
>
> (Donald Trump, citing Mussolini, 28 February 2016, Twitter)

When Chuck Todd, host of "Meet the Press" on NBC, questioned Donald Trump about this tweet and the quote, Trump answered: "Look, Mussolini was Mussolini. It's okay to — it's a very good quote, it's a very interesting quote, and I know it. I saw it. I saw what — and I

know who said it. But what difference does it make whether it's Mussolini or somebody else? It's certainly a very interesting quote." Contrary to Trump's argument, speaker attributes can influence the interpretation of their statements. Imagine a poor person telling you that they own a boat, then imagine a rich person telling you that they own a boat - you probably just imagined two different boats. Just as that boat, many, if not all expressions and descriptions are ambiguous to some degree. Every object has an infinite number of features. Thus, by default, descriptions are incomplete. Moreover, descriptions are communication and as such underlie maxims of relevance and conciseness (Grice, 1975), which require omission. Therefore, when a Christian or Muslim is described as "religious", or when a Buddhist or CEO describes herself as "rich" or "faithful", this leaves quite some room for interpretation and may challenge our meaning-making processes more than we think. The act of interpreting encountered ambiguities in line with social stereotypes could also reinforce our perception of those around us in terms of stereotypes and categories: The retrogressive Muslim and the Christian with the protestant work ethic, the wise Buddhist and the boastful CEO. Stereotypes like these are very persistent, and seemingly "confirmed" very quickly. This often happens even when both stereotyped and stereotyping individuals are aware of the issue and try not to perpetuate them in conversations and actions (Maass, 1999). Interpreting ambiguous information in a stereotype-consistent way might be a process that can subvert these efforts in order to reinforce stereotypes and categorization. Such a feedback loop of re-stereotyping would neither depend on nor produce any visible cues and would be concealed by our persistent overestimation of our own perceptual accuracy (Duncan, 1976). Stereotyping and categorization are closely intertwined. Categories are cognitive sorting units, while stereotypes are their associated attributes (or sorting criteria), used both to identify category exemplars and to mutually assimilate exemplars of the same category in perception. While the effect of stereotype-consistent interpretation on stereotype reinforcement seems more

straightforward (Slusher & Anderson, 1987), its effect on reinforcing categorization could have consequences that reach beyond the perceiver's perception of the target person. Only by means of categorization can the effects of stereotype-consistent interpretations be generalized to third others within the same category. Thus, we hypothesize that a statement does not have to contain stereotype-consistent information to trigger re-stereotyping of a person and thus carry forward the perception of that person in terms of a category. An information gap that can be "imputed" with stereotype-consistent content by the perceiver should suffice to reinforce categorization.

3.2.1 Construal bias and stereotype imputation

That imputing stereotypes into ambiguous information about individuals results in biased construal of these individuals (also: "meaning change", Asch, 1952; Bryson & Franco, 1976; "imaginal confirmation", Slusher & Anderson, 1987; "stereotypes color trait and behavior ratings in the presence of ambiguous information", Kunda & Thagard, 1996; "stereotypes cue interpretation of speaker intent", Pexman & Olineck, 2002; "stereotypes influence the interpretation of ambiguous social behavior", Sagar & Schofield, 1980) has been demonstrated across diverse context domains. Regarding performance evaluation, (randomly allocated) academic test results were considered better when they ostensibly came from a child with higher vs. lower socioeconomic background (Darley & Gross, 1983), or when children had common, popular, and attractive names rather than rare, unpopular, and unattractive names (Harari & McDavid, 1973). Depending on whether a person that "hit someone annoying" was dubbed a construction worker or housewife, the action was interpreted as "punching an adult" or "spanking a child" (Kunda & Sherman-Williams, 1993), and Black and white children perceived the same ambiguous shove as more violent when it was carried out by a Black rather than a White person (Sagar & Schofield, 1980). Participants' elaborations of the same situation contained more details consistent with

stereotypes about a fictional character's occupation (e.g. (wealthy) lawyer – expensive car) than the originally presented situation, and more stereotype-consistent situations were imagined "around" fictional characters beyond the presented situations (Slusher & Anderson, 1987). The term "revolution" in a political quote by Thomas Jefferson was understood more like "agitation" (59%) than "revolution" (1%) when it was attributed to himself, but understood more like "revolution" (68%) than "agitation" (9%) when attributed to Lenin (Asch, 1952). Examining a special case of ambiguity, Pexman and Olineck (2002) found that the same statement was more likely to be understood as sarcastic when uttered by a person with a "sarcastic occupation" like comedian, talk show host or movie critic rather than a "less sarcastic occupation" like army sergeant, accountant or doctor. Thus, construal bias has been found on ambiguous behavior (Darley & Gross, 1983; Harari & McDavid, 1973; Kunda & Sherman-Williams, 1993), traits (Bryson & Franco, 1976; Slusher & Anderson, 1987), and statements (Asch, 1952; Pexman & Olineck, 2002).

Commonly discussed consequences of construal bias on the side of the perceiver include shifting standards in perceiving target persons and their actions (Kunda & Sherman-Williams, 1993) and improving future ability to detect intentions (of ironic intent, Pexman & Olineck, 2002). In an exception to the predominantly theoretical discussion of effects on individual target persons, Slusher and Anderson (1987) found empirical evidence for an effect of construal bias on stereotype maintenance. Participants were given three types of situation descriptions: stereotype-confirming, stereotype-irrelevant and stereotype-relevant situations. As they imputed stereotypes into the stereotype-relevant (generation) situations and forgot they were self-generated, participants subsequently overestimated the percentage of stereotype-confirming adjectives in the stimuli they had seen before. Whereas this study provides some evidence for the actual process of imputation and its role in stereotype maintenance, in the present project we aim to go one step further: Imputing stereotypes into

statements uttered by various speakers should also reinforce categorization. If an individual is imagined as more stereotype-consistent than is supported by the information ecology, this *imaginal confirmation* (Slusher & Anderson, 1987) should make the associated category appear more relevant to the perception of that individual. This should increase the perception of that individual in terms of the category. In contrast to effects of construal bias affecting only target persons, if construal bias increases the perception of an individual in terms of the category (i.e. categorization strength) and the category therefore becomes more salient, it could be increasingly applied to similar individuals. This way, stereotype maintenance could generalize to unrelated category exemplars. Therefore, we aim to follow the effect back to the level of social categorization.

3.2.2 Category reinforcement vs. reconstructive category guessing

Taking the previous logic to the empirical field of social categorization would allow the prediction of more intra- than inter-category errors in a "Who said what?" Paradigm when statement construability is high, while the asymmetry between intra- and inter-category errors should be less pronounced when statement construability is lower. To recap, a typical WSW consists of two phases: an encoding phase ("discussion phase"), in which 8 "speakers", 4 from each category (e.g. Germans and Syrians), are presented sequentially paired with statements. Each speaker is presented 6 times, and every trial features a new statement, resulting in 48 subsequent presentations of binary speaker-statement pairs. In the surprise recall phase ("assignment phase"), all statements, plus as many new statements, are presented again. Participants must choose for each statement, which one of the speakers "said" it – or whether it was not presented previously. When participants are confronted with a statement which they cannot reallocate to the correct speaker, they may use a speaker category attribute as proxy to increase their chance at guessing the correct speaker (S. E. Taylor et al., 1978). For example, they might not remember that Ahmad said it, but that the speaker's name

sounded Arabic. In this case, the participant should randomly choose a speaker from the correct category for their answer – resulting in more within-category errors. This is traditionally assessed by the error-difference measure that compares the sums of within- and between-category errors (S. E. Taylor et al., 1978). A higher within- than between-category error rate is attributed to the application of social categories in the WSW task. Construal bias in the WSW would take place already in the discussion phase. In the logic of the WSW paradigm, this would then lead to a higher perception and memory of speakers in terms of their category, increasing the error-difference-measure of social categorization. We would take this as an example of construal bias increasing categorization proper. Importantly, however, construal bias could also influence the WSW results in an identical way by reconstructive category guessing (based on imputed information) in the assignment phase. We will argue below why we believe that both possibilities are conceivable and informative, but before doing so, we will briefly explain the second reconstructive option in detail.

When the stereotype-relevant statements presented in the discussion phase are presented again in the assignment phase, participants could remember the stereotypes they associated with the statements earlier in the discussion phase. The thereby enriched statements would then point to the correct category on their own. This way, it would become unnecessary to remember attributes of the speaker of a statement (from the discussion phase) for indicating the correct speaker category. For example, when a speaker stated that they owned a boat in the discussion phase, in the assignment phase, the participant could remember imagining a fishing boat before. Therefore, she could judge the statement to be stereotype-consistent with the "poor" category and assign the statement randomly among the poor speakers. This would increase the error-difference-measure in the same way as category memory. However, using the imputed stereotype as a retrieval cue for the speaker category in this manner does not require the attachment of statement content to the speaker in the discussion phase, as

originally intended in the WSW paradigm to measure category memory. Arguably, as the statement therefore stops being the neutral, irrelevant link between the perceived speaker in the discussion phase and the recalled speaker in the assignment phase, categorization in the assignment phase is not unprompted and spontaneous anymore. This is reminiscent of "reconstructive category guessing", a process identified as confounding and inflating measured categorization in the WSW paradigm (Klauer et al., 2014; Klauer & Ehrenberg, 2005). (Expectancy-based) reconstructive category guessing "relies on stereotypical pre-experimental expectancies about the likely category origin of a given kind of statement. In expectancy-based guessing, the statement is assigned to a member of the category that is stereotypically associated with the statement content. […] Because it relies on features stereotypically associated with the category, it is an instance of applying categorical knowledge." (Klauer & Ehrenberg, 2005, p. 499). This definition refers to statements that already contain stereotype-consistent information. Construable stereotype-relevant statements, however, would require multiple applications of categorical knowledge to enable reconstructive category guessing. In addition to the connection between statement-stereotype and speaker-category that has to be established, the stereotypes extracted from the statements to improve category guessing in the assignment phase must have been imputed from the speaker category in the discussion phase earlier on. Thus, if a derivative of reconstructive category guessing is indeed the underlying process through which construal bias influences the error-difference measure in the WSW paradigm, construal bias might indeed not lead to increased category memory. The error-difference measure (or parameter d) might, however, still reflect multiple instances of applying categorical knowledge and possibly category reinforcement. For example, imagine someone you knew to be poor told you about their boat one day, which made you think about a fishing boat. When you meet that person again a week later, they might have to mention their boat again (as a retrieval cue) for you to remember the

fishing boat and reinforce your perception of that person as poor. However, you will now believe you remember two instances of that person telling you about their fishing boat, still solidifying your perception of that person as "poor person". We therefore argue that both suggested ways in which construal bias might influence the WSW categorization measures are relevant to categorization, address the research question in a valid way, and are capable to produce novel insights into the relationship between construal bias and social categorization. Another issue of a more technical nature is the measurement of construal bias.

3.2.3 Measuring construal bias

Measuring construal bias poses a substantial challenge. Social categorization can be seen as the starting and end point of stereotype imputation as a process: Firstly, target persons must be perceived in terms of a category membership for stereotypes to become accessible for imputation. Secondly, construal bias should then turn a stereotype-relevant statement into a stereotype-consistent one in perception, reinforcing the perception of the target person in terms of the category. Unfortunately, when investigating construal bias, social categorization is the only unidimensional concept and common denominator across participants and stimuli in this circular process. The process of stereotype imputation in the narrower sense between these two instances of categorization, however, is highly idiosyncratic and dependent on stimulus content. When a poor man states that he has a boat, perceivers are free to imagine a fishing boat (implying that man has a job that is not well-paid), or a houseboat in London (because he can't afford a flat there), or a self-built model of a boat using leftover wood (signifying a creative hobby instead of a job). Each of these manifestations of construal bias creates a poor-rich distinction on another stereotype dimension: Low-paid vs. well-paid job, precarious vs. comfortable housing conditions, unemployed vs. employed. A single interpretation in line with one of these stereotype dimensions can even point towards multiple second-tier stereotype dimensions and corresponding inferences (low-paid job: lazy-active,

untalented-talented, idealist-pragmatist), while each different stimulus (fishing boat, car, downtown flat) suggests another set of stereotype dimensions on its own. To reiterate, all these examples portray valid instantiations of construal bias. Thus, any one stereotype dimension chosen to measure construal bias across as many stimuli and participants as possible may still fail to grasp the extent of construal bias sufficiently, decreasing power and measurement validity. Earlier research on this phenomenon already grappled with this problem and came to similar solutions to the ones we applied in our pretests: open responses. Yet, open responses requesting descriptions that specify an ambiguous object or situation (e.g. car: Rolls Royce, truck, family car…) must usually be back-rated manually and for each item individually. This becomes resource-intensive when manipulating construal bias not only within a single statement, but across a large pool of statements. In the present studies, we manipulated construal bias within the statements of the WSW discussion phase. We assumed that along the criteria mentioned in Section 4.1, statements differ in the degree to which they lend themselves to stereotype imputation regarding a given category dimension. As, probably, all statements are construable to some degree, for each of the following studies, we designed and pretested two statement pools: one was more prone to stereotype imputation (or "construable") than the other regarding the target category dimension.

In the studies reported here, we therefore introduced a proxy to stereotype imputation as manipulation check. We reasoned that oftentimes, differentially imputed meaning would also lead to evaluations of statements that ultimately differ between speaker categories. For example, in the context of Syrian refugees fleeing to Germany, a Syrian or a German could both comment that they worry a lot. Self-stereotypes of Germans include a propensity to worry a lot about minor issues. Even if not, the worries of a Syrian refugees could be interpreted in a much more serious way: Maybe they have family members that are threatened or that they lost, financial and other existential problems. Because of these differential

interpretations, the unspecific "worries" expressed by the Syrian refugee might elicit more comprehension or understanding by a participant than the same feeling expressed by a German. Thus, we used differential *evaluation* as proxy for construal bias. After each presented statement, we asked participants, how much compassion they felt towards the statement made by the respective speaker.

We are aware that this measure does not measure construal bias directly, but through the indirect route of evaluation of the imputed meaning. This introduces additional measurement and inferential fuzziness. Differential imputed meaning might not create differential evaluation, and differential evaluation could be based solely on the evaluation of the speaker category (e.g. appreciating the achievements a female politician (categorically) more because she is a woman, not because (stereotypically,) women have it harder in politics). While the former case (construal bias without differential evaluation) seems plausible but would (only) lead to an inflated beta error and thus lower power, the latter case (differential evaluation without construal bias) seems less plausible. The "blind" evaluation of someone's statements based solely on their category membership would require actively suppressing the interpretation and evaluation of those statements, a behavior that humans seem to engage in intuitively (Sherman et al., 2011). Thus, we apply the differential evaluation measure as a compromise, and aware of its shortcomings.

The present research

To investigate whether construal bias reinforces social categorization, we altered the WSW paradigm to accommodate stereotype imputation. We demonstrate construal bias in three studies across two category dimensions, nationality (Studies 3.1 and 3.2), and age (Study 3.3). In Study 3.1, we used a statement evaluation task as encoding phase in a signal detection paradigm to study the effects of construal bias. The statement evaluation task served as a manipulation check for the difference in instigated construal bias between conditions, and

reinforced stereotype imputation. In Study 3.2, we implemented a full WSW paradigm. The results confirmed our findings from Study 3.1 and enabled us to observe the effects on social categorization strength directly. In Study 3.3, we generalized the effect to the age category dimension. Statement pools for Studies 3.2 and 3.3 (high and low construal statements) were pretested in separate pre-studies. They indicated the presence of a construal bias and differential construability between statement sets (see online supplement). This research line aims to provide new insights into a subtle mechanism of meaning-making during impression formation that might contribute to the maintenance of stereotypes and categorization. We report all studies in this research line, except two. One provided mixed evidence as construal bias correlated with categorization, but the manipulation of statement ambiguity did not affect social categorization experimentally. The other did not show an effect of manipulation either, likely due to a deviating manipulation design (see online supplement). We report all measures, manipulations, and exclusions in these studies. Final sample size was determined before data collection. All materials, data and supplemental analyses are available on our OSF project site (https://osf.io/vzcne).

3.2.4 Study 3.1

In Study 3.1, we primarily aimed to get a first empirical grip on the concept of construal bias. Thus, we used a signal detection paradigm less meticulous but likely more robust than the WSW paradigm. We focused on manipulating statement construability and set the construal bias evaluation measure as dependent variable for our confirmatory hypothesis in the preregistration. Yet, the data collected within this study can still hint at a possible connection between construal bias and social categorization. We used the signal detection measure d' as exploratory proxy for social categorization. The higher the hit rate (frequency of German statements recognized as German) and the lower the false alarm rate (frequency of

Syrian statements falsely considered as German), the better the detection of the signal, measured by sensitivity index d'. We chose ethnicity as category dimension for this study. We suspected that differences in stereotypical representations between ethnicities would pervade nearly all domains of life – particularly those of self-identified "native Germans" about their own group in contrast to Syrian migrants represented by typically Muslim names (see Duncan, 1976, for a similar approach). Secondly, to match statements across conditions as closely as possible while maximizing the difference in construability, statements were constructed to be identical across conditions, with the exception that highly construal statements were self-referential or referred to the speaker ingroups. Lastly, we chose only the 40 statement pairs with the best pretest results from a pretest candidate pool of 200. They were paired with four speakers (two from each category), in a source memory task without (new) distractor statements. We predicted that statement evaluations would differ more strongly between speaker categories for the ambiguous, highly construable statements than for the less construable statements. Furthermore, a higher sensitivity (d') for the high construal bias condition would be in line with our overall hypothesis regarding category reinforcement. Study 4.1 was preregistered at https://aspredicted.org/blind.php?x=5dt4ip.

Method

Participants. As preregistered, $N = 60$ students took part in the lab at the University of Cologne, Germany in exchange for €5. The data of all participants (25 men, 35 women, $M_{age} = 23.87$, $SD_{age} = 7.10$) were included in the analysis. We report here post-hoc sensitivity analyses for achieved power in the evaluation measure. The ANOVA with two within factors had 80% power to detect an effect size of $\eta_p^2 = .06$ in the present study.

Stimuli/ Manipulation. Ten common German (e.g. "Christian", "Hannes", "Alexander") and 10 common Arab names (e.g. "Mohammad", "Ahmad", "Tarek") comprised the speaker pools, from which 2 names were drawn respectively for each participant anew. No two names

sounded or were spelled similarly, in order to reduce non-categorical speaker confusion. Forty statement-pairs were chosen via a pretest (see online supplement). They were selected for maximum difference in construability between conditions as rated by 4 student assistants who were blind towards the hypothesis but not the construct of construal bias itself. Raters were instructed that they would be shown several statements made either by a person who grew up in Germany or by a person who had asylum seeker status in Germany at the time. They were then asked to estimate the probability of the stated person saying the stated sentence, i.e. how well the statement fitted the person on a scale from 1[not likely / does not fit] to 10 [very likely / does not fit]. Each of the statements was rated by each rater, but only in relation to one of the categories, resulting in two ratings for each statement on each category. Statement-pairs included e.g. “We are not wanted in this country. / Some people think that they are unwanted in this country.”, “Discrimination against us is a big problem in Germany. / Discrimination is still a big problem in Germany.”. For each statement-pair, we then computed the difference in construability between them. The ideal pair would consist of a high construal statement that fit perfectly to both a German and a Syrian speaker (due to the statement being maximally prone to stereotype imputation from either category), and a low construal statement that fit both categories minimally. Moreover, we did not select statements that fit much more to one category over the other, as this would indicate stereotype-consistency instead of construability. A statement-pair that fit these requirements well was e.g. “The family has a completely different meaning in our society. / The family has a great significance in all societies.”. As the statements were mainly complaints about current circumstances of speakers’ current lives, and the living situation of Syrian refugees in Germany was still far from ideal at the time, we also assumed that a construal bias should result in more understanding for statements when made by ostensibly Syrian speakers. For exploratory reasons, we also included a measure of political orientation commonly used in Germany, the

"Sunday Question": Participants were asked to indicate which party they would vote for if there was a general election the following Sunday.

Procedure. Participants took part in the study in the lab at University of Cologne, Germany, where they gave informed consent and performed a signal detection task. They were instructed that they were about to see "comments from panelists discussing the effects of the flight movement to Germany as a result of the Syrian war." Then, the participants were presented with successive paired presentations consisting of a speaker and a statement each. Statement condition was varied within subject, so that all participants saw all statements. Statements were randomly assigned to speakers irrespective of category membership. The construal bias measure was integrated into the discussion phase in this study: After every speaker-statement pair presentation, participants were asked to indicate how much they understood that the speaker made the respective statement. There was no inter-trial break before the next stimulus pair was presented. Then, participants moved on to the surprise recall task. In it, all statements from the presentation phase (80) were shown in random order, and participants were asked "Who said that?" each time. They responded by ticking one of four speakers / answer options. Then the participants were asked to indicate their age, gender and political orientation, and were debriefed.

Results and Discussion

As preregistered, a two-way ANOVA was conducted that examined the effect of (statement) condition and speaker category on statement evaluation. We found the predicted interaction between (statement) condition and speaker category on statement evaluation, $F(1, 59) = 15.59$, $p < .001$, $\eta_p^2 = .21$, in that statement evaluation differed more between speaker categories in the high construal condition than in the low construal condition, confirming our main hypothesis. Additionally, there was a main effect of condition: Less construable statements received more understanding by the participants across speaker categories, $F(1,$

59) = 15.59, p <.001, η_p^2 = .21 (Fig. 3.1). For the signal detection analysis, answers were collapsed across group membership: Hits (selection of a German speaker for a statement made by a German speaker), Misses (selection of a Syrian speaker for a statement made by a German speaker, False Alarms (selection of a German speaker for a statement made by a Syrian speaker), and Correct Rejections (selection of a Syrian speaker for a statement made by a Syrian speaker). Although the frequency of selecting a correct individual speaker could have been computed based on the data to get a measure of individuation, participants had a chance of 50% of guessing the correct speaker, so this measure would not have been particularly meaningful. On the signal detection sensitivity / source memory measure, we found an effect in the presumed direction (high construal condition: $M_{d'}$ = .47, $SD_{d'}$ = .38; low construal condition: $M_{d'}$ = .22, $SD_{d'}$ = .36; $t(58) = 4.02$, p <.001, d_z = .54, *95% CI*$_d$ [.17, .91]; ROC-curve see Fig. 3.2).

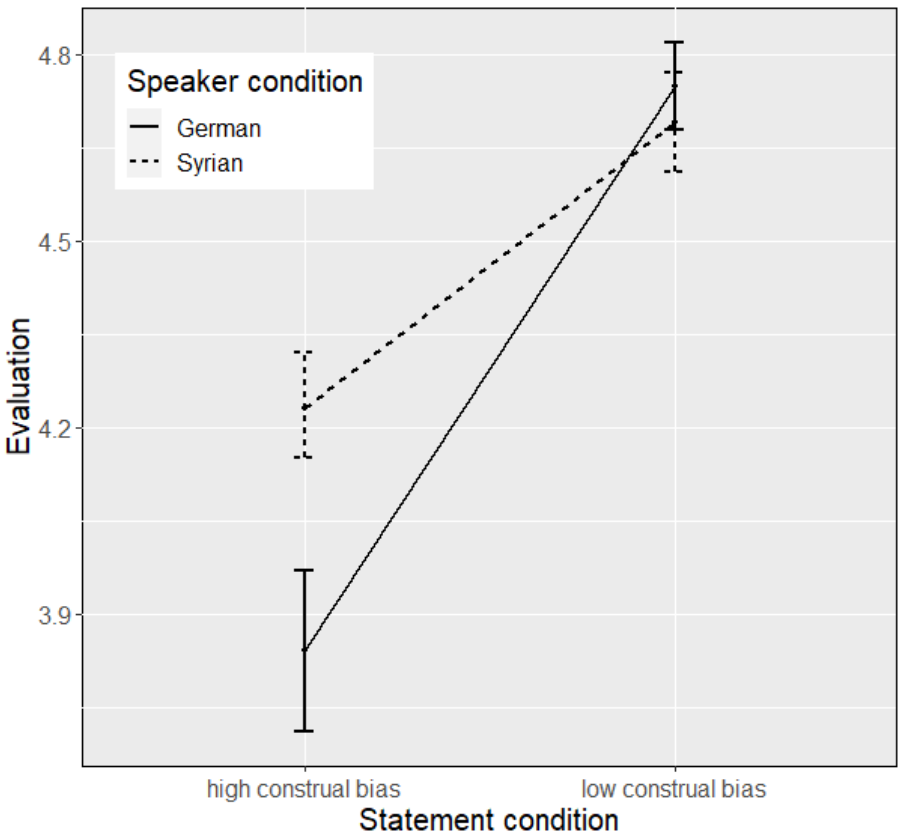

Figure 3.1. Evaluation measure (means and 95% CIs) in Study 3.1.

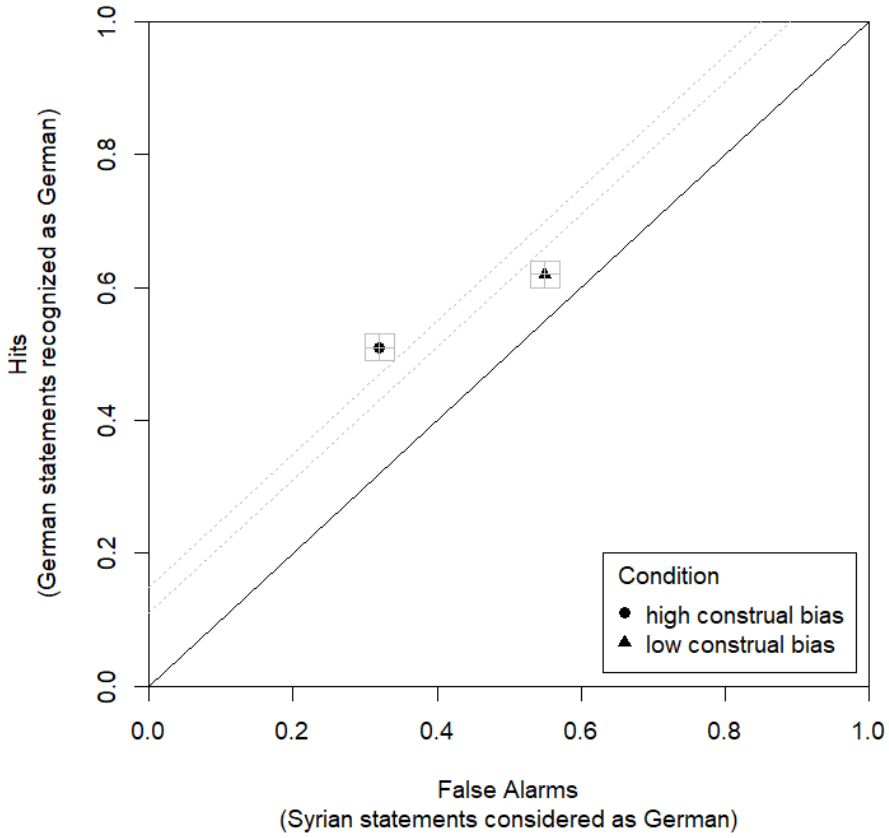

Figure 3.2. Receiver operating characteristic (Mean d' and 95% CIs) for Study 3.1. Sensitivity for statements uttered by German and Syrian speakers is plotted as a function of Hits (frequency of German statements recognized as German) and False alarms (frequency of Syrian statements falsely considered as German). The drawn through diagonal denotes performance that is not different from chance: There are as many Hits as False alarms. The area under the diagonal (indicating negative performance) is thus typically not needed. The opposite diagonal denotes sensitivity: The more Hits and the fewer False alarms, the better recognition performance in this task. As can be seen from the point estimates (auxiliary lines along the confidence intervals for better readability), memory is more sensitive for highly construable statements than for less construable statements (falling diagonal). Participants also had a more conservative response criterion (rising diagonal) towards less construable statements.

Sensitivity was therefore higher for highly construable statements. Participants also had a more conservative response criterion (Fig. 3.2, rising diagonal) towards less construable statements, meaning that these statements were generally more often attributed to Germans than to Syrians.

The results supported the hypotheses, in that statement evaluation differed more between speaker categories when statements invited construal bias. Moreover, high construability also led to increased source memory sensitivity for the statements. Two auxiliary findings might be explained by additional differences of statement content between conditions. The less construable statements often called for improvement for all people or people suffering from a shortcoming irrespective of category membership, while the highly construable statements often demand exclusive treatment for the own group. Thus, the former might have elicited much more understanding in the participants. If this is true, the response bias towards attributing low construal statements to German speakers might reflect a pro-ingroup bias: "inclusive" and "reconciling" statements are expected more from (German) ingroup members than from (Syrian) outgroup members. Having established stimuli that vary in construability successfully manipulate source memory, we aimed to target social categorization more specifically in the next study.

3.2.5 Study 3.2

To investigate whether construal bias can reinforce categorization, in Study 3.2, we used an adapted WSW paradigm. This allowed us to distinguish between different mental processes contributing to the WSW categorization measure. The within-subjects design was retained to increase power, such that each participant saw 24 low construal and 24 high construal statements in the discussion phase. We predicted, in line with Study 3.1, that categorization denoted by the d parameter should be higher in the high construal condition. In line with Study 3.1, we also predicted that evaluations should differ more strongly between speaker categories for the high construal bias statements. Study 3.2 was preregistered at https://aspredicted.org/blind.php?x=jr24a5.

Method

Participants. As preregistered, $N = 100$ students took part in the lab at the University of Cologne, Germany in exchange for €5. The data of all participants (41 men, 55 women, 4 other, $M_{age} = 23.03$, $SD_{age} = 5.00$) were included in the analysis. The classical error-difference measure had 80% power to detect an effect size of $d_z = .28$ in the present study.

Stimuli/ Manipulation. Speaker and statement stimuli were the same as in Study 3.1, 16 statements were added to the stimulus set to complete the statement pool (see online supplement).

Procedure. Same as in Study 3.1, but featuring 4 speakers per category, as well as 48 instead of 80 statements in the discussion phase, while another 48 statements served as distractors. Statements were equally distributed across speakers, speaker categories, conditions and targets vs. distractors. They were randomly distributed across all cells (except across conditions).

Results and Discussion

The predicted interaction between (statement) condition and speaker category on statement evaluation was significant, $F(1, 99) = 31.97$, $p < .001$, $\eta_p^2 = .24$, in that statement evaluation differed more between speaker categories in the high construal condition than in the low construal condition. Additionally, there were main effects of condition (less construable statements received more understanding by the participants across speaker categories, $F(1, 99) = 205.10$, $p < .001$, $\eta_p^2 = .67$) and speaker category (statements by Syrian speakers were preferred to speakers from German speakers, $F(1, 99) = 22.78$, $p < .001$, $\eta_p^2 = .19$).

The error-difference-measure indicated higher categorization in the high construal condition (high construal bias: $M = 4.67$, $SD = 3.94$; low construal bias: $M = 1.81$, $SD = 4.32$), $t(99) = 5.93$, $p < .001$, $d_z = .62$, 95% - CI [.34 - .91]. However, we preregistered the MPT

analysis as relevant analysis, given appropriate model fit. Model fit was appropriate when equality-restricting d_a and d_b in both conditions (high construal bias: $T_1^{observed} = 0.154$, $T_1^{predicted} = 0.077$, $p = .07$, $T_2^{observed} = 4.24$, $T_2^{predicted} = 4.89$, $p = .60$; low construal bias: $T_1^{observed} = 0.114$, $T_1^{predicted} = 0.077$, $p = .23$, $T_2^{observed} = 6.77$, $T_2^{predicted} = 4.54$, $p = .12$).

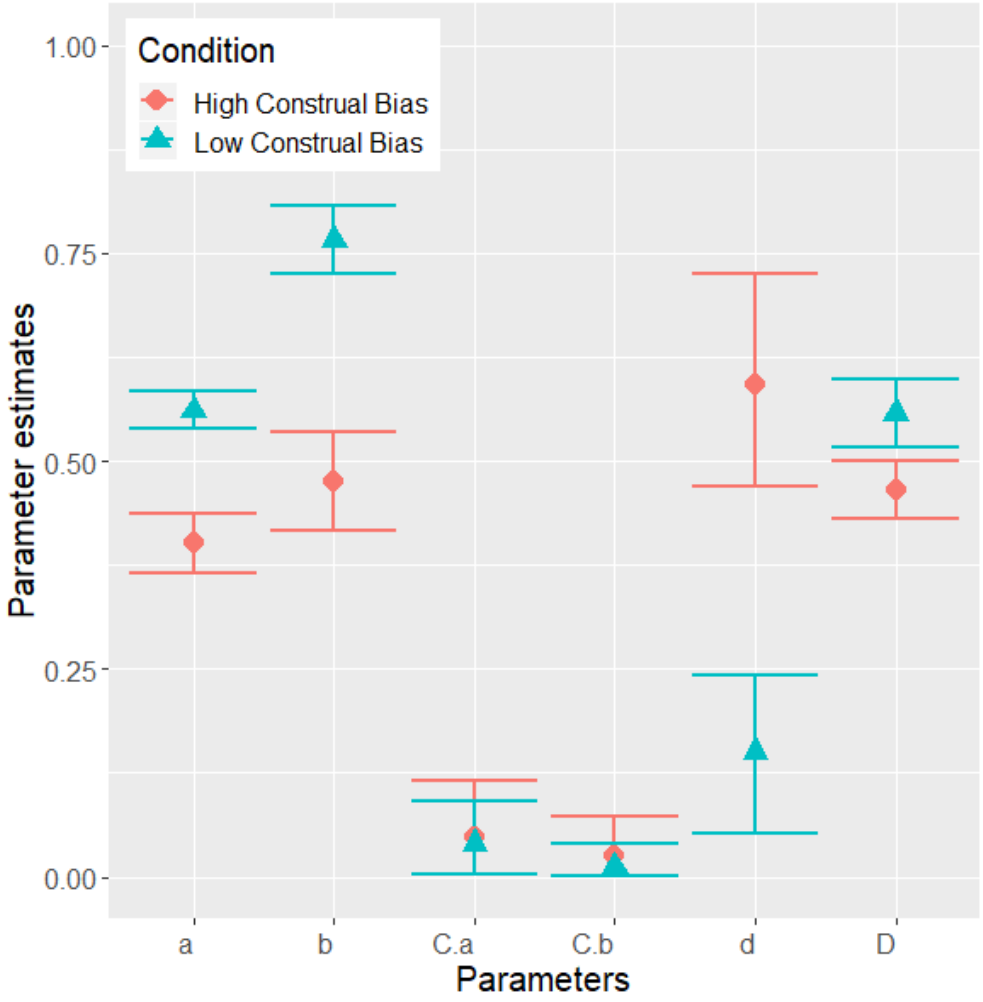

Figure 3.3. Mean parameter estimates and 95% CIs by condition in Study 3.2. The model distinguishes between statement memory (parameter *D*), exclusive person memory (parameter *C*), category memory (parameter *d*) and biases in category- or old-new (expectancy-based) guessing (parameters *a, b*). Parameters *C* (person memory) and *d* (categorization) can be computed for both target categories separately. Subscript X_a denotes the German category, whereas subscript X_b denotes the Syrian category.

Table 3.1
Parameter Estimates and 95% CIs in Study 3.2

	High construal		Low construal		
Parameter	*M*	95% CI	*M*	95% CI	p_B
a	.40	[.37, .44]	.56	[.54, .58]	<.001*
b	.48	[.42, .53]	.77	[.72, .81]	<.001*
C_a	.05	[.00, .12]	.04	[.00, .09]	.43
C_b	.03	[.00, .07]	.01	[.00, .04]	.28
d	.59	[.47, .73]	.15	[.05, .24]	**<.001***
D	.47	[.43, .50]	.56	[.52, .60]	<.001*

Note. Mean parameter estimates and 95% Credible Intervals in Study 3.2. Applied restrictions within conditions: $d_a = d_b$, $D_a = D_b = D_n$. Bayesian *p*-value (p_B) for difference in parameter estimates between conditions.

The comparison between categorization parameters across conditions confirmed the abovementioned results - categorization strength was significantly higher in the high construal condition ($\Delta d = .44$ with the 95% credibility interval [.29, .60], $p_B < .001$, Fig. 3.3). Furthermore, seeing low construal statements was associated with increased guessing-based attribution of the statement to the ingroup ($p_B < .001$), increased guessing of a statement to having been seen before ($p_B < .001$), and increased statement detection ($p_B < .001$, see Table 3.1). Additionally, the strength of construal bias (as measured by the difference in statement evaluation between categories in the high construal condition) correlated positively and significantly with categorization strength (as measured by error-difference) both in the high construal condition ($r(99) = .29$, $p = .004$) and the low construal condition ($r(99) = .33$, $p = .001$). Statement evaluation in the low construal condition did not correlate with any measures of social categorization.

The results supported the hypotheses in that statement evaluation differed more between speaker categories when statements were prone to construal bias. Moreover, high construability also led to increased social categorization of the speakers. As in Study 3.1, less construable statements were judged more favorably. In Study 3.2, the main effect of speaker category also became significant, in that statements made by Syrian speakers were preferred. Both error-difference analysis and MPT results converged in finding that high construal

statements increased categorization of their speakers relative to low construal statements. The manipulation, however, did not have an exclusive effect on categorization. In line with the response bias observed in Study 3.1, low construal statements were more often attributed to the ingroup. These statements might have also seemed more similar to participants, as much more new statements were treated as previously seen in the low construal condition. This could even be a side-effect of construal bias: If presented high-construal statements were imputed with stereotypes (i.e., group attributes), this could have made them more distinct from new statements (and mutually more dissimilar), leading to more correct guessing of high-construal statements. Somewhat in conflict with the effect of condition on categorization strength and guessing of speaker category (parameter b), less construable statements were correctly recognized as new or belonging to the correct category slightly better. On one hand, such significant differences on parameter D sometimes emerge unsuspectedly and might not carry meaning (Klauer & Wegener, 1998). On the other hand, less construable statements might indeed be easier to recognize in general, as the lack of construal bias leads to maximal overlap between presented and remembered statement, i.e. familiarity (parameter D), while biases have a larger impact on the statements not remembered (parameter b). Although the correlations between evaluation and categorization support the notion of shared variance that could be attributed to construal bias, the lack of specificity regarding the effects triggered by the manipulation suggests possible alternative explanations. For instance, the highly construable statements could have been more emotionally charged overall, e.g. because they generally elicited more negative valence. The political context of an immigration peak in Germany and referring to possible points of conflict between the two social groups may have influenced the study in a similar manner as envisaged for construal bias and introduced dangers to internal validity. Thus, we designed the next study to generalize the effect to age, another fundamental social category dimension.

3.2.6 Study 3.3

In order to generalize the effect to another category dimension and therefore disperse some of the alternative hypotheses mentioned above, we transferred the effect to the age category dimension ("old" vs. "young"). While the setting still portrayed a potential intergroup conflict over status and material resources, we expected that its effects would not be as pervasive and unspecific as in the ethnicity setting. Thus, we again predicted effects of condition (statement construability) on differential statement evaluation and social categorization strength. Study 4.3 was preregistered at https://aspredicted.org/blind.php?x=dn5gy4.

Method

Participants. As preregistered, data collection ended on the day the 100th dataset was collected. $N = 107$ students took part in the lab at the University of Cologne, Germany in exchange for €5. The data of all participants (40 men, 65 women, 2 other, $M_{age} = 23.20$, $SD_{age} = 5.61$) were included in the analysis. The classical error-difference measure had 80% power to detect an effect size of $d_z = .27$ in the present study.

Stimuli/ Manipulation. The youngest and oldest white male faces were chosen as speaker portraits from the Face Database (Minear & Park, 2004, see online supplement for selection of stimuli). Statement sets for high and low construal conditions were compiled from scratch but adhering to the same construction principles as the statement sets in Studies 4.1 and 4.2 (see online supplement). Statement-pairs included e.g. "My generation respects people who earned that respect. / People respect people who they think earned that respect.", "Women's roles are very different in my generation. / Women have different roles in different generations."

Procedure. Same as in Study 4.2.

Results and Discussion

The predicted interaction between (statement) condition and speaker category on statement evaluation did not become significant, $F(1, 106) = 0.14$, $p = .71$, $\eta_p^2 = .001$, so statement evaluation did not differ more between speaker categories in the high construal condition than in the low construal condition. The main effect of condition became significant again (less construable statements elicited more understanding in the participants across speaker categories, $F(1, 106) = 172.11$, $p < .001$, $\eta_p^2 = .62$).

The error-difference-measure indicated a difference in categorization between conditions (high construal bias: $M = 6.40$, $SD = 3.63$; low construal bias: $M = 2.92$, $SD = 3.97$), $t(106) = 6.65$, $p < .001$, $d_z = .67$, 95% - CI [.40 - .95]. However, we again preregistered the MPT analysis as relevant analysis, given appropriate model fit. Model fit was appropriate when letting d_a and d_b free to vary in the high construal condition. To compare conditions analytically, we let the parameters vary freely in both conditions (high construal bias: $T_1^{observed} = 0.151$, $T_1^{predicted} = 0.072$, $p = .06$, $T_2^{observed} = 5.19$, $T_2^{predicted} = 4.77$, $p = .44$; low construal bias: $T_1^{observed} = .069$, $T_1^{predicted} = 0.072$, $p = .50$, $T_2^{observed} = 4.11$, $T_2^{predicted} = 4.65$, $p = .63$). There was a difference between conditions on categorization strength, both for young speakers ($\Delta d_a = .55$ with the 95% credibility interval [.23, .86], $p_B = .001$) and old speakers ($\Delta d_b = .72$ with the 95% credibility interval [.50, .94], $p_B < .001$, Fig. 3.4). As can be seen in Figure 3.4, differences between other parameters were much lower than in Study 3.2. Nevertheless, low construal statements were still more often guessed to having been said by a young / ingroup speaker (parameter a; $p_B = .009$), and were slightly more often guessed to be seen in the discussion phase (parameter b; $p_B = .04$), and person memory was slightly better for young speakers in the high construal condition (parameter C_a; $p_B = .007$), while statement memory was slightly better in the low construal condition (parameter D; $p_B = .05$, see Table 3.2). In summary, we succeeded in finding the predicted effect of statement construability on

categorization strength. We did not find the predicted effect on statement evaluation. So far, we have treated differential statement evaluation as manipulation check for the difference in statement construability between conditions. Indeed, it may be an efficient way to capture a major dimension of differential perceptual processing of ambiguous information by speaker category when that difference is present. Yet, it is important to keep in mind that evaluation is just one dimension of differential appraisal of ambiguous information, and that it may not be relevant in all settings. In fact, as previously mentioned, between each participant and statement can emerge unique sets of semantic stereotype dimensions, and their overlap may be difficult to design and capture. We predicted that the intergenerational conflict setting would produce a similar differential evaluation as the ethnic intergroup setting in Studies 3.1 and 3.2. However, the statements apparently elicited a similar degree of positive appraisal and understanding whether they were presented as made by a young or old speaker. This does not mean that no difference in construal bias was present between conditions – but taken together with the smaller, but still significant effects of condition on other parameters, this does not eliminate the danger of alternative hypotheses. The most obvious one might be connected to the category (self-)references included in the high construal statements to make them construable. Simply making the speaker category salient in that way might straightforwardly lead to more categorical encoding and, therefore, recall. Although not predicted, the lack of evaluation polarization in the high construal condition could clear another alternative explanation. In the previous studies, the forced statement evaluation in the discussion phase could have externally prompted the differential construal, on the basis of which categorization could have been reinforced not spontaneously but based on a double priming. This could not have been the case in the present study.

Table 3.2
Parameter Estimates and 95% CIs in Study 3.3

	High construal		Low construal		
Parameter	*M*	95% CI	*M*	95% CI	p_B
a	.48	[.45, .51]	.54	[.50, .57]	.009
b	.60	[.55, .65]	.66	[.61, .71]	.04
C_a	.26	[.17, .34]	.13	[.07, .19]	.007
C_b	.14	[.05, .21]	.06	[.00, .12]	.08
d_a	.90	[.74, 1.0]	.35	[.07, .63]	**<.001***
d_b	.46	[.43, .50]	.51	[.47, .56]	**<.001***
D	.99	[.94, 1.0]	.27	[.04, .49]	0.05

Note. Mean parameter estimates and 95% Credible Intervals in Study 3.3. Applied restrictions within conditions: $d_a = d_b$, $D_a = D_b = D_n$. Bayesian *p*-value (p_B) for difference in parameter estimates between conditions.

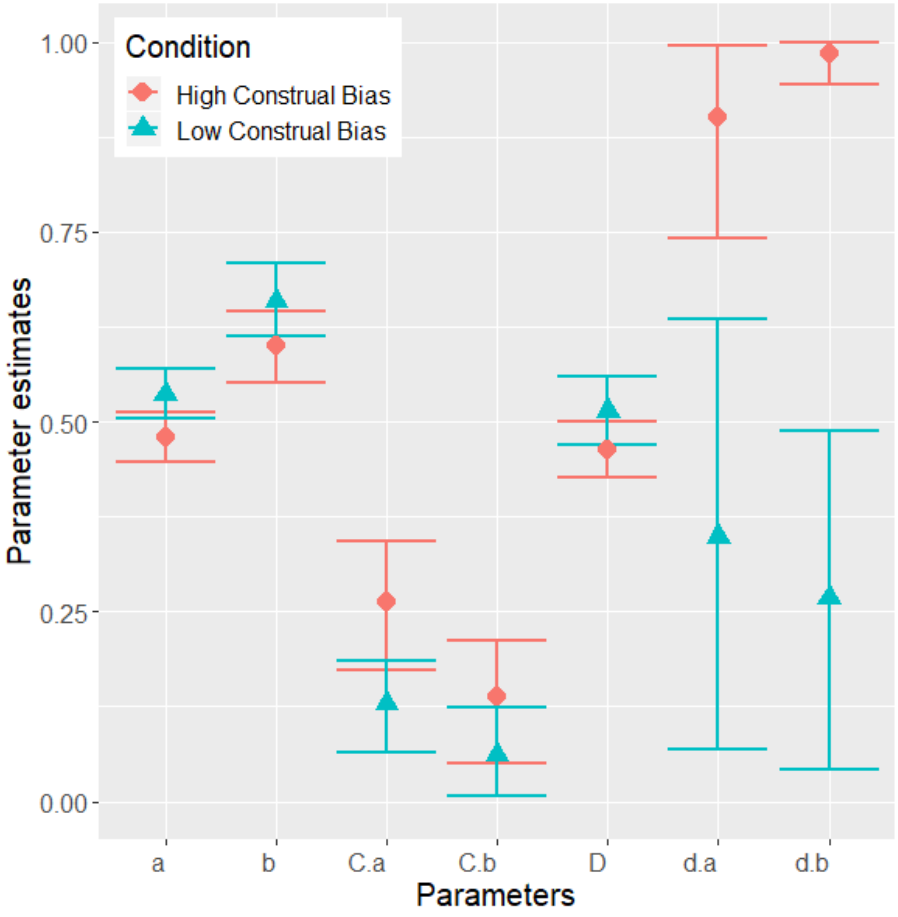

Figure 3.4. Mean parameter estimates and 95% CIs by condition in Study 3.3. The model distinguishes between statement memory (parameter *D*), exclusive person memory (parameter *C*), category memory (parameter *d*) and biases in category- or old-new (expectancy-based) guessing (parameters *a, b*). Parameters *C* (person memory) and *d* (categorization) can be computed for both target categories separately. Subscript X_a denotes the young category, whereas subscript X_b denotes the old category.

3.2.7 General Discussion

In three studies, we established that construal bias supports category reinforcement. We generalized the effect across two category dimensions (nationality / ethnicity and age) and two paradigms (signal detection, WSW). To this aim, we focused on only one category (dimension) at a time throughout this line of research, which has been criticized before (e.g. research on intersectionality, Bodenhausen & Peery, 2009). Yet, our approach enabled us to let the stereotype content activated and applied in the studies vary freely between participants and broadly between stimuli. This adds to the previous literature, which often focused on only one stereotype dimension, or very few of them (e.g. aggression along the race dimension, hireability along the gender dimension). Thus, we both (re-)established construal bias as a concept and portrayed its relevance by applying it to social categorization. Besides phenomena like self-fulfilling prophecies (Rosenthal & Jacobson, 1968) or spontaneous trait inferences (Uleman et al., 2008), construal bias might therefore be another subtle unobtrusive mechanism of everyday stereotype perpetuation that is based on memory and recall biases. The presented studies did not shed light on whether construal bias primarily influences category memory / social categorization proper or reconstructive category guessing. Which of the two processes underlie the effect reported in the previous studies, remains an empirical question.

3.2.7.1 Retracing the vicious circle between categorization and stereotyping

Indeed, the concept of construal bias might be most relevant in its projected role in the "vicious circle" of re-stereotyping and re-categorization. Between categorization and stereotyping, the scientifically more established direction of causality leads from categorization to stereotyping (e.g. Bhatia, 2017; Rees, Ma, & Sherman, 2020). It seems trivial: Categories can exist without stereotypes, but stereotypes, as category-based schemata, cannot exist in the absence of categories. This is supported by research on the influence of

cognitive load on categorization and stereotyping: Categorization requires neither attentional nor motivational resources, while stereotype activation requires at least one of them (Sherman et al., 2011). Stereotype application, on the contrary, occurs under lack of attentional resources – if they are available, however, individuation or stereotype suppression / inhibition can take place. Yet, this does not preclude stereotypes from reinforcing categorization, which can then lead to a spreading in stereotypical perception to other target person features (e.g. visual features, Durante & Fiske, 2017; Hugenberg & Bodenhausen, 2004; MacLin & Malpass, 2001). It would be interesting to locate construal bias within these dynamics. By definition, construal bias requires both stereotype activation and application and leads to biased perception of an initially unbiased information ecology. While it seems relatively automatic and effortless in theory, it remains an open question whether stereotype imputation is similarly effortful or motivation dependent as other cognitive processes related to stereotyping. In summary, in the studies presented, we laid out the effect of stereotype imputation on generalized stereotype maintenance halfway. Construal bias does reinforce social categorization. It remains to be seen, however, whether construal bias-instigated category reinforcement can lead to recategorization and restereotyping generalizing of unrelated category exemplars.

Besides its value for understanding processes of stereotype and categorization maintenance, construal bias might also help to better understand social categorization itself. If the “lumping and splitting” of others - that is social categorization (Zerubavel, 1996) - includes meaningful variance over and above simple interindividual similarity of target persons (as found in the WSW paradigm, Degner et al., 2020), construal bias could account for part of the variance that differentiates between people but is not dependent on the present information ecology. After all, stereotype imputation should also be possible in one-on-one encounters with target persons. In a reversal of the categorization-stereotyping causal chain

(and etiology) favored above, stereotype imputation could also play a major role in the emergence of social categorization in infants. Imagine someone inviting a child to play. The kind of play activities to be expected may vary considerably by the prospective playmate's age, role (parent, parents' friends), and maybe even gender. This would make it functional to expect different play activities depending on a playmate's category (Bigler & Liben, 2006, 2007; Liberman, Woodward, & Kinzler, 2017). Subsequently, categories for which construal bias proves functional may stabilize as cognitive schemata, ready to be applied to the next ambiguous information ecology.

Beyond social categorization, combined with the Linguistic Intergroup Bias (Maass, 1999) construal bias could also reinforce intergroup bias. As negative outgroup descriptions and positive ingroup descriptions are often formulated more abstractly than their counterparts, they could facilitate more stereotype imputation for more extreme negative outgroup stereotypes and positive ingroup stereotypes respectively, leading to an even steeper intergroup evaluation gradient. Similarly, combined with the findings of Chapter 2, an initially asymmetrical categorical encoding may result in us remembering more individual features of ingroup speakers, but outgroup speakers more as category exemplars. This may favor construal bias for outgroup statements an further perpetuate intergroup bias.

Lastly, construal bias could also contribute to understanding unintended processes underlying the WSW paradigm. If construal bias indeed leads to reconstructive category guessing, this may possibly challenge the construct validity of the WSW paradigm. If all statements are construable towards speaker categories and stereotypes to some degree, construal bias could inflate the measure of social categorization in any study using the WSW paradigm.

3.2.7.2 Further limitations and future directions

In the present studies, we used evaluation ratings as construal bias proxy that we expected to differ between categories, with mixed success. Establishing a more efficient and/or reliable measure of stereotype imputation could facilitate future research on this phenomenon immensely. Our current measures of construal bias could have also enforced stereotype imputation that would not have occurred spontaneously, which could have led to an overestimation of its effect on social categorization. In failing to use a stereotype dimension relevant to the category dimension in Study 3.3, however, we might have produced counterevidence against this objection.

On a more conceptual level, auxiliary assumptions and boundary conditions of construal bias still need to be established. Abstractness of expressions is likely to be positively related to stereotype imputation, but it remains to be seen if it is a linear relationship or rather an exponential one. One more level of abstractness in the Linguistic Category Model (Semin & Fiedler, 1991) might equal multiple additional interpretations and subsumed categories. One step removed, abstractness of expressions might be related to categorization strength in an inverted u-shape: Maximally abstract expressions might not activate stereotype dimensions as easily as expressions of intermediate abstractness, as the former may violate principles of accessibility and fit. It would also be interesting to see whether any expressions approach "non-construability", and which range of construability is inherent in everyday communication. This might help understand the real-life relevance of construal bias in the formation and maintenance of categories and stereotypes.

Much research has been devoted to the perception of stereotype-consistent and stereotype-inconsistent information (Sherman et al., 2011). We believe that information that is merely stereotype-relevant without leaning to either side is as important to consider and completes the classification of external information types that can trigger stereotyping. To integrate the

concept into the established theoretical structures, future research needs to investigate the influence of cognitive load on stereotype imputation and construal bias, as well as the relative precedence of ambiguous stereotype-relevant information and stereotype-(in-)consistent information in social perception and meaning-making. For example, in earlier research, subjects also appeared to neglect stereotypes when unambiguous individuating information was offered next to it (Kunda & Sherman-Williams, 1993).

Lastly, construal bias is a social psychological phenomenon rooted in linguistic "matter". From a linguistic standpoint, in the present paper, we merely focused on the "semantic" content level and did not consider illocutionary (expressive) and perlocutionary (appellative) aspects of pragmatics. Whether they influence and proliferate construal bias in a qualitatively or quantitatively similar manner remains an open question.

We found first evidence that construal bias reinforces social categorization. Therefore, as another variant of a psychological "vicious circle", construal bias could be one of the reasons why people might find it difficult to shake their shadows – or stop casting them onto others. Understanding construal bias could relieve some of the irritation that comes along with us realizing that our mind sometimes develops a momentum of its own – and maybe give us a chance to outsmart it anyway.

Chapter 4 – Decategorization under common threat

Social categorization strength is often deemed robust towards external influence and contextual change. After examining a possible contributing mechanism for this in the last chapter, in the present chapter, we study the malleability of social categorization – by a common threat. Chapter 4 consists of two sections. In Section 4.1, we establish the effect of a common threat on decategorization, in Section 4.2, we propose three processes that might explain this effect.

4.1 Unite Against: A common threat invokes spontaneous decategorization between social categories

Abstract

A frequent rhetoric in the political arena calls members of larger groups like nations to lay aside all dividing differences and unite in face of a common threat. In the present research we sought to test whether such a unifying effect of external threat already manifests in such basic cognitive processes as automatic categorization even for such strong schisms as the ones between black and white Americans or Israeli Jews and Arabs. In Studies 4.1 & 4.2 (*N*=183/144, USA), we established the decategorization effect in the context of black and white US Americans. In Study 4.3, we showed the effect again in a German lab for the gender category (*N*=101). In Study 4.4 (*N*=168, Israel), we transferred the effect to the context of the Israeli-Palestinian conflict and teased apart the separate effects of intergroup threat, common goal and common threat, and category membership of participants. In summary, a "common enemy" leads to the decategorization of social groups already at an early automatic stage.

Human history can be seen as a history of divisions such as wars, conflicts and discrimination. But at a closer look, there also seem to be antagonistic forces driving people together (again): For example, NATO originated from a common fear of western European countries and the USA of a possible military soviet aggression, and in an attempt to settle these differences in return, Mikhail Gorbachev reportedly agreed with Ronald Reagan, who suggested to pause the Cold War in the case of an "alien invasion" (Orr, 2009). These are just two of the more ostensive examples of a figure of political action and argumentation that is referred to in addresses by head of states all around the globe: a "common enemy", against which formerly opposing groups unite. In the present paper, we delineate and test the novel hypothesis that such a uniting effect of a common threat does not just happen out of strategic considerations but manifests itself at a very early and automatic level of intergroup cognition: social categorization.

4.1.1 "Common enemy" and common threat in intergroup research

Research on intergroup attitudes is built around the notion that the social environment is categorized into distinct entities (i.e., social groups) that are then associated with attributes (stereotypes) and valence (prejudice). For these later processes, it has been well established that a common threat ameliorates intergroup prejudice and increases liking. Effects of an experimentally induced common enemy or threat were found on intergroup liking between newly established groups or groups with low everyday salience (Sherif & Sherif, 1953) and on prejudice towards "basic" social groups, i.e. national groups (Adachi, Hodson, Willoughby, & Zanette, 2015) or races (Feshbach & Singer, 1957). A common threat also increased intergroup helping (Batson et al., 1979; Dovidio & Morris, 1975; Hayden, Jackson, & Guydish, 1984; J. B. Taylor, Zurcher, & Key, 1970; van Leeuwen & Zagefka, 2017), and cooperation (Greitemeyer, Traut-Mattausch, & Osswald, 2012; Pepitone & Kleiner, 1957; Wright, 1943). All the effects mentioned are situated in ingroup-outgroup settings and thus

mainly rest on social identification as motivational basis for observed intergroup dynamics. Also, these effects are either evaluative (liking, prejudice) or more downstream behavioral outcomes (helping, cooperation). Yet, threat may trigger more basal spontaneous reactions, so that a common threat could affect earlier psychological processes of social categorization.

Before one can attach a presumed attribute or a valence to social categories, one needs to construct or "see" them. Social categorization processes reduce the overwhelming complexity of the social environment by lumping humans into neat categories based on race (Blacks, Whites), gender (women, men), age (adults, children), ethnicity or other more or less arbitrary dimensions. Importantly, these processes are theoretically independent of social identification (Rosch, 1978) and highly automatic. While categorization along such dimensions is deeply engrained in human socialization and is thus usually highly salient, it is also situationally malleable, e.g. by introducing competing categories (Klauer et al., 2014). In the present research we sought to test the novel idea that such perceived categorical divides can be attenuated if the two implied categories face a common enemy. This goes beyond previous existing research on spontaneous social categorization by extending categorical malleability over and above direct perceptual competition. Regarding the common enemy effect, the present research investigates whether a common threat reduces categorization, without necessarily implying other antecedents like being personally affected. Consider two distinct groups (A) and (B), which are both targeted by a common threat (C). In much research on the common enemy effect, perceived unification was measured in participants that were explicitly addressed by the setup as members of either group A or B, thus, they were personally affected by the threat (this applies to both intergroup threat research, Greitemeyer et al., 2012; Pepitone & Kleiner, 1957, and the Robbers Cave Study, Sherif & Sherif, 1953). Although a common threat may bring members of A and B closer to each other, is it currently not known

whether such salient self-identification and thus becoming a direct target of the common threat is indeed necessary to perceive the unification of groups A and B.

Also, in previous research, the "common threat" manipulation often included a concerted common effort or action of the target groups towards and/or against the threat, such as cooperation (Adachi, Hodson, Willoughby, & Zanette, 2014; Greitemeyer et al., 2012). It was thus unclear whether the observed effect of increased cooperation was an outcome of confrontation with a common threat (Wright, 1943; Pepitone & Kleiner, 1957; Greitemeyer, Traut-Mattausch, & Osswald, 2012) or just the result of directly activating the notion of cooperation, thus bordering circularity: Cooperation against a common enemy was found to increase cooperation in a subsequent task (Greitemeyer et al., 2012).

Thirdly, a superordinate ingroup was often offered to the participants, prompting them e.g. to indicate how American they felt after the manipulation. This is in line with the widely held theoretical assumption that social categorization cannot dissolve, but only dissipate between more or less inclusive levels of categories (Rosch, 1978; Sherif & Sherif, 1953), which suggests that only recategorization on a superordinate level can attenuate categorization on a lower level (Brewer & Miller, 1984; Drury et al., 2009; Gaertner et al., 2000). As this assumption was never tested to the knowledge of the authors, it remains unclear whether recategorization is a necessary precondition for the common enemy effect.

Lastly, the "common enemy" is a construct that has become deeply embedded in social discourse and representation. Therefore, study participants may hold similar popular lay beliefs regarding effects of a common enemy and may be more sensitive to detecting the construct. This may increase their sensitivity to demand characteristics in experimental settings (Sharpe & Whelton, 2016). For example, lay beliefs about the ability to exert self-control have been shown to influence measured self-control (Job, Dweck, & Walton, 2010) and belief in a fixed human nature is associated with dynamics in intergroup bias (Hong et al.,

2004). Therefore, it is vital to reduce demand characteristics to a minimum when studying common enemy effects. To measure social categorization strength, we use a more unobtrusive paradigm that is based on performance rather than response preference, and is thus less susceptible to such demands.

Another challenge are imprecise definitions and operationalizations that can be found regarding the outcome variable(s) in common threat studies. Several studies find that common threat increases "social cohesion" (Greitemeyer et al., 2012; Wright, 1943). However, the definition of this concept is still debated in psychology (Bruhn, 2009). Sometimes it refers to similar characteristics within one's group (Deutsch, 1968), and at other times includes emotional components (e.g. "feelings of cohesion" as "feeling a bond", Greitemeyer et al., 2012). This moves the concept closer to (intergroup) liking again. Similarly, cooperation as an effect of common threat (Greitemeyer et al., 2012; Wright, 1943) does not necessarily imply decategorization of the cooperation partners. We argue that these assumptions at the levels of independent and dependent variables are not necessary and that common threat can lead to decategorization in a perceptual, spontaneous manner.

4.1.2 Common threat as unifier in social categorization

Evidence for a uniting power of common threat at the early stage of social categorization can be found in diverse studies and theoretical frameworks. Firstly, common threat could unify members of distinct social categories by recategorization (Brewer & Miller, 1984; Drury et al., 2009; Feshbach & Singer, 1957; Gaertner et al., 2000; Vezzali, Drury, Versari, & Cadamuro, 2016). Specifically, a common enemy could make cross-cutting category memberships salient or enhance the salience of features shared by categories (Gaertner et al., 2000). The Social Identity Model of Collective Resilience (SIMCR, Drury et al., 2009; Vezzali et al., 2016) suggests that common threat could lead to the experience of a common fate that could lead to a shared social identity, activating shared goals (Drury et al., 2009).

Furthermore, a common enemy enhances intergroup liking (Sherif & Sherif, 1953) and prejudice reduction (Burnstein & McRae, 1962; Feshbach & Singer, 1957). As liking can breed similarity (Cartwright & Harary, 1956; Collisson & Howell, 2014; Heider, 1946), a common enemy could also lead to decategorization by increasing intergroup liking. Taken together, these lines of theorizing and empirical findings converge in the prediction that a common threat should reduce the tendency to distinctively categorize category members along category boundaries. Thus, we hypothesize that introducing a common threat reduces categorization strength in a classic paradigm of automatic spontaneous categorization: the "Who Said What?"-Paradigm.

The present research

To address the question whether a common threat reduces social categorization, we conducted a research program out of which we report four studies (see below for the studies not included). In these studies, we tested whether spontaneous social categorization is reduced in the context of a common enemy (compared to a neutral baseline). We interpreted categorization as spontaneous as the target categories were never mentioned. In all studies, the manipulation was induced by statement sets that differed between conditions, while statements were randomly assigned to speakers within conditions (Klauer et al., 2014). To approach the unique contribution of a common threat to decategorization, the common threat condition was compared to neutral baseline conditions (Studies 4.1, 4.3, and 4.4), an intergroup threat context (Studies 4.2 and 4.4) and a common goal context (Study 4.4). To establish generalizability across cultural contexts, two of the studies were conducted in the US (MTurk) and one was conducted in Germany and Israel respectively (both in the lab). Target categories were chosen that were salient in each of the cultural contexts respectively (e.g. Israeli Jews and Arabs in Study 4.4) to maximize external validity and increase generalizability across target categories. Likewise, "discussion" topics (statement content)

was matched to the cultural context. While in the US and Germany, Islamist terrorist threat was chosen as a realistic threat, we switched to "disease" in the Israel study (see Study 4.4 for details). Thus, we were also able to generalize from "human" common enemies to "non-human" common threats. To increase generalizability not only on participant level but also on stimulus level, in Studies 4.2 - 4.4 the four speaker stimuli (portraits in Studies 4.1 and 4.2, names in Studies 4.3 and 4.4) were sampled from sets of 7-10 in both categories. Also, none of the studies share the exact same statement set(s). We report all measures, manipulations, and exclusions in these studies. Final sample size was determined before data collection. Upon completion, no further data was collected. The 4 studies reported here are part of a research program that spanned 11 studies in total. For reasons of brevity, we focus on the most relevant studies in the current paper, but detailed information on all studies can be found in the online supplementary materials. Importantly, although not all studies yielded significant support for our focal hypothesis, a meta-analysis across all studies yielded a robust effect (see meta-analytic integration below). All materials, data and supplemental analyses are available on our OSF project site (https://osf.io/urw3h).

4.1.3 Study 4.1

As an initial test of the hypothesis that a common threat reduces the perceived group boundaries already at the early stage of spontaneous social categorization, we conducted a study on race categorization among US citizens on Amazon MTurk. As a "common enemy", we chose a threat very salient and persistent in US public discourse: Islamist terrorism (Kearns, Betus, & Lemieux, 2019; Sui et al., 2017). We hypothesized that the category boundaries between Black and White US Americans could soften in the face of Islamist terrorism, a threat towards all US citizens. Translated to the parameter specification of the

MPT Model for the WSW task, we expected an exclusive reduction of the d parameter estimate indicating categorization strength when the common threat was made salient.

Method

Participants. One hundred and ninety-eight US-Americans took part in the study on Amazon Mechanical Turk in exchange for $4. An automatic filter only allowed them to participate if they had not participated in any previous WSW study conducted by the authors' lab. If participants indicated at the end of the study that they either saw their data not fit for analysis ($n_{threat} = 1$, $n_{neutral} = 3$) or that they had taken notes during the experiment ($n_{threat} = 3$, $n_{neutral} = 8$), their data were not analyzed. These two exclusion criteria were the only ones used for all studies. Thus, the data of 183 participants ($n_{threat} = 99$, $n_{neutral} = 84$, 91 men, 92 women, M_{age}=34.09, SD_{age}=10.21, 146 White, 8 Hispanic/Latino, 17 Black/African American, 2 Native American/ American Indian, 11 Asian/ Pacific Islander, 5 other, 4 did not wish to answer) were included in the analysis. Power analysis is not yet available for the hierarchical Bayesian implementation of MPT models in the R package 'TreeBUGS' (Heck et al., 2018), therefore, we determined a-priori sample sizes by compromising between the current standard in the social categorization literature using MPT analysis and new standards in the field of social psychology. Thus, we report here post-hoc sensitivity analyses for achieved power in the error-difference measure. The classical error-difference measure had 80% power to detect an effect size of $\eta_p^2 = .02$ in the present study.

Stimuli/ Manipulation. Four portraits of white Americans and four portraits of black Americans were chosen from the Chicago Face Database (Ma et al., 2015). The portraits displayed faces with a neutral expression which scored high exclusively on the white vs. black race pre-rating in the CFD coding manual. More specifically, Black portraits were rated by a very high percentage of raters to be black (99%) but by no one to be white. Likewise, White portraits were rated by 95% to be white but by no one to be black. Both sets of pictures

differed in no other rating available for this set (age, self-rated maleness, self-rated femaleness, Asian, Latino, multi-ethnicity, other ethnicity, fear, anger, attractiveness, Babyface, disgust, dominance, femininity, happiness, masculinity, prototypicality, sadness, suitability, surprise, threat, trustworthiness, unusualness). To implement the manipulation in a between-design, we designed two sets of statements: The baseline set, adapted from Klapper et al. (2016), contained statements about housing, hobbies and work (e.g. “I need to travel to work an hour every day”, “Usually, I go to work by train”, “I live in a rented house with balcony.”): The statement set in the common threat condition introduced Islamist threat to the US (e.g. “Every Islamist can find bomb recipes online.”, “Thankfully, many Islamist organizations hate each other.”, “Drones do not defend America against the Taliban.”). See online supplementary material for complete statement set.

Procedure. After accepting the HIT, participants accessed the study via a link to the study on the SoSci Survey platform (Leiner, 2014), where they gave informed consent and indicated their gender and age, and performed the WSW task, randomly assigned to one of two conditions: Common (Islamist) Threat or Neutral. They were instructed that they were about to see several “young people meeting for the first time and engaging in a dialogue”. Then, the participants were presented with successive paired presentations consisting of a speaker and a statement each. Statements were randomly assigned to speakers irrespective of category membership. The speaker was presented first and for 9 s, while the statement was displayed after a 1.5 s delay, so both stimuli were then simultaneously displayed for 7.5 s. There was no inter-trial break before the next stimulus pair was presented. After observing all 48 pairings, participants moved on to the surprise recall task. In the surprise recall task all statements from the presentation phase (48) and distractor set (48, in total 96 statements) were shown in random order, and participants were asked "Who said that?" each time. They responded by

ticking one of nine answer options, namely the eight portraits and the option "None. This statement is new."

Afterwards, they answered a manipulation check question about their perceived level of threat (“How threatened did you feel?”). For exploratory purposes, participants then answered the Symbolic Racism 2000 Scale (Henry & Sears, 2002), Multiculturalism and Colorblindness scales (A. Hahn, Banchefsky, Park, & Judd, 2015), the Similarity Focus scale (Ohmann & Burgmer, 2016), the Assessment of Ingroup-Outgroup overlap (Aron, Aron, & Smollan, 1992; Schubert & Otten, 2002) and a measure of political orientation on a visual analogue scale from liberal to conservative. Then the participants were debriefed and asked to indicate their age, gender and ethnicity as well as questions about their perceived data quality and whether they took notes during the study.

Results and Discussion

Although for all studies, the primary analysis was the one via MPT model, we will also report the classic frequentist analysis by within-between ANOVA with errortype (intra-vs. inter-category) as within- and experimental condition as between-subjects factor (S. E. Taylor et al., 1978), for ease of understanding and readability. However, all tables, graphs and interpretations offered are based on the MPT results. Perceived threat differed between conditions, in that participants felt more threatened in the common threat (M = 32.81, SD = 29.73) than in the neutral (M = 10.45, SD = 20.42) condition (t (173.62) = 5.99, p <.001, d = 0.86). Speaking to the hypothesis that the extent of automatic categorization (i.e. more within- than between category errors) was contingent on the experimental manipulation, there was a significant interaction of error type and condition. Indicating reduced categorization under threat, the difference between intra- and inter-category errors was significantly lower in the common threat condition (M = -0.82, SD = 6.13) than in the neutral control condition (M = 2.49, SD = 6.25), F(1,181) = 13.07, p < .001, η_p^2 = .07.

We then computed Bayesian hierarchical latent-trait MPT models (Klauer, 2010) by means of the R package 'TreeBUGS' (version 1.2.0, Heck et al., 2018) for both conditions separately. In addition to the restrictions suggested by Klauer and Wegener (1998), category memory parameters d_a and d_b were restricted to be equal within conditions, as there were no a-priori hypotheses regarding a difference between the two (see also Klauer et al., 2014). Model fit was appropriate in both conditions (Common Threat: $T_1^{observed} = 0.086$, $T_1^{predicted} = 0.075$, $p = .39$, $T_2^{observed} = 11.21$, $T_2^{predicted} = 13.25$, $p = .62$; Neutral: $T_1^{observed} = 0.102$, $T_1^{predicted} = 0.088$, $p = .37$, $T_2^{observed} = 12.63$, $T_2^{predicted} = 13.49$, $p = .55$).

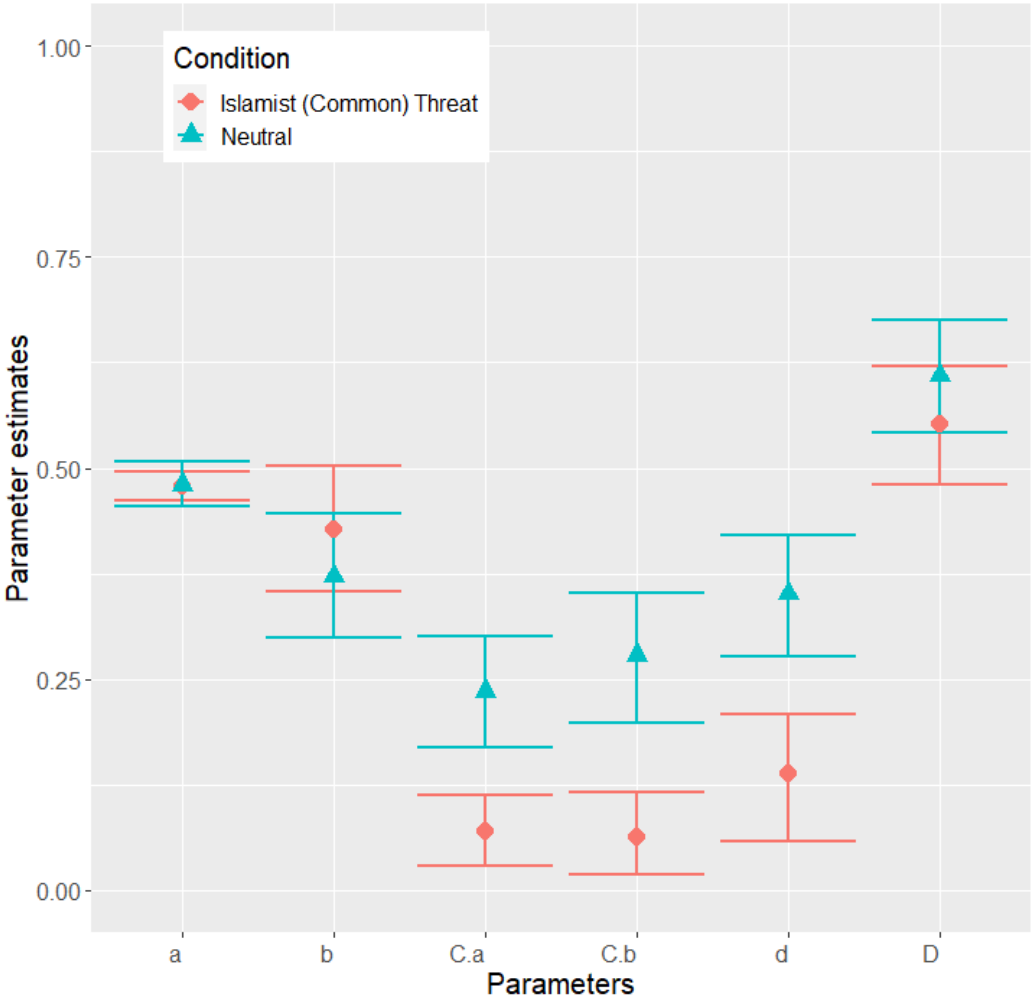

Figure 4.1. Parameter estimates and 95% CIs by conditions in Study 4.1.

Table 4.1
Parameter Estimates and 95% CIs in Study 4.1

	Common Threat		Neutral		
Parameter	*M*	95% CI	*M*	95% CI	p_B
a	.48	[.46, .50]	.48	[.46, .51]	.43
b	.43	[.35, .50]	.37	[.30, .45]	.85
C_a	.07	[.03, .11]	.24	[.17, .30]	< .001*
C_b	.06	[.02, .12]	.28	[.20, .34]	< .001*
d	.14	[.06, .21]	.35	[.28, .42]	**< .001***
D	.55	[.48, .62]	.61	[.54, .68]	.11

Note. Mean parameter estimates and 95% Credible Intervals in Study 1. Applied restrictions within conditions: $d_a=d_b$, $D_a=D_b=D_n$. Bayesian *p*-value (p_B) for difference in parameter estimates between conditions.

As hypothesized, there was a significant reduction in category memory in the threat condition compared to the neutral condition (Δd =.21 with the 95% credibility interval [.11, .32], $p_B < .001$, Fig. 4.1). There were also effects of manipulation on person memory in the same direction (see Table 4.1). In principle, it is conceivable that people categorized less because they were – in light of threat – more alert overall, increasing memory for individual speakers and thus decreasing the likelihood of answers based on memory for category attributes only. The fact that the person memory parameters (C_a, C_b) did not increase but decrease speaks against this interpretation. On the other hand, individual person memory decreasing at the same rate as categorization strength could indicate an overall shift of attention to the statements as source of threatening information, away from the speakers and their individual and category attributes. As there was a slight trend in this direction, we continued to monitor the C parameter estimates in later studies. There were no significant differences between any other parameter estimates between conditions (see Table 4.1), and no significant correlations between the categorization strength parameter estimate and any of the additional measures (see online supplementary material).

As hypothesized, the manipulation had an exclusive effect on categorization strength, in that the context of a common threat reduced categorization between black and white US

Americans. Both classical and MPT analyses converged regarding the result. Despite this, some caution seems warranted. Categorization is inferred from disproportionately attributing statements for which there is no correct memory of the speaker to another speaker from the same category (rather than a speaker form the opposite category). For this to happen, participants need to at least rudimentarily make the statement-speaker connection (regarding speaker category). Statement attributes, however, differed substantially between conditions, not only as intended regarding their threatening capacity, but also in their length and complexity of both grammar and content. Thus, differences in semantic complexity between conditions could potentially lead to more attention to the statement, but less to the speaker stimulus it was presented with (and thus less memory also for speaker category). Also, as the threat statements could simply be more interesting to the subject, the threatening capacity of the statement content could lure the subjects' attention away from the speaker stimulus (and the implied category). Moreover, the neutral statements were more about the individual speakers' lives, so it might have been more natural for subjects to try and memorize the information as connected to the speakers than in the threat condition. Although this explanation would be in line with the decreasing tendency in person memory between conditions, it is not clear why this effect would be boosted for the categorization parameter (compared to the person memory parameter). Nevertheless, it seemed advisable to prepare more comparable statement sets for the follow-up studies.

Furthermore, the WSW paradigm at least partly operationalizes categorization as similarity between persons within a category relative to between persons across categories. This makes it vulnerable to too strong similarities in a subgroup of a category. Imagine Jack, Jake, William and George engaging in a discussion. As Jack and Jake are very similar, they are confused more, leading to an increased d parameter. But not because both of them are white or male (i.e. are both perceived to belong to the target category), but because their

names look and sound very similar. Again, this would not influence the decategorization effect, but overall categorization could be inflated. We thus aimed at avoiding such confounds of category with inter-exemplar similarity by sampling category exemplars from a larger stimulus pool in the following studies.

4.1.4 Study 4.2

In Study 4.2, the statement sets for the control condition were compiled from scratch and the ones for the experimental condition were heavily revised and adapted, to form two sets in which each statement in one condition had a syntactic twin in the other condition. We aimed at providing two statement sets that were highly comparable in their semantics, length, notation and valence. By controlling for differences in syntactic and semantic complexity between conditions in such a design, differential categorization strength can no longer be attributed to idiosyncrasies of specific statements or statement sets but can only evolve from the contextual meaning they transport. The statement sets only differed in whether the discussed threat targeted both categories from the outside (common enemy/ common threat) or constituted a threat between them (mutual/ intergroup threat). The manipulation was implemented by exchanging “Islamists” (Common Threat) with “Racists” (Intergroup Threat). As we retained the “realistic” setup of these statements, a few minor adjustments had to be made to some statements’ content, too. Importantly, to keep likelihood of a statement being said by a speaker of either category about equal, “racist” statements were, if possible, worded in an ambiguous way. Thus, “We need more protection against violent outbursts from members of the other race.” would express Anti-Black criticism of a white US American, but Anti-White criticism of a black US American, creating an intergroup threat context. In this and the following studies, we also randomly sampled the four speaker portraits from a pool of

8 for each participant anew. The study was preregistered (http://aspredicted.org/blind.php?x=bg48b6).

Method

The study followed the identical procedure as Study 4.1 with a few alterations explained below.

Participants. One hundred and forty-nine US Americans completed the study on Mturk. N=5 were excluded because they either indicated that they had taken notes in the discussion phase ($n_{control}$=1) or they did not see their data fit for analysis (n_{threat} = 2, $n_{control}$ = 2). Of the 144 participants in the final dataset, 53.5% were male, 45.1% female, 1.4% other, M_{age}=33.08, SD_{age}=8.94, n_{threat} = 75, $n_{control}$ = 69. Participants received $3 for their participation. The classical error-difference measure had 80% power to detect an effect size of η_p^2 = .03 in the present study.

Stimuli/ Manipulation. Eight portraits of white Americans and eight portraits of black Americans were chosen from the California Face Database. The portraits displayed faces with a neutral expression which scored high exclusively on the white vs. black race pre-rating in the CFD coding manual. More specifically, Black portraits were consistently rated to be black (100%) and not white (0%). Likewise, White portraits were rated by 100% to be white but by no one to be black. Both sets of pictures differed in no other rating (see Study 4.1 for list of ratings). Statements were constructed as outlined above, resulting in stimuli paired across conditions such as: “Islamist ideologies fuel many current wars.”, “Racist ideologies fuel many current unrests.”; “Drones can’t protect us from Islamism.”, “Guns can’t protect us from racism.”; “I watched a video where an Islamist cut a US captive’s throat.”, “I watched a video where a racist beat up a man with my skin color.”. Statement list in the online supplementary material.

Results and Discussion

As in this study the control condition portrayed intergroup threat, we did not expect a difference on our manipulation check question on perceived threat. Perceived threat did not differ between common threat (M = 70.59, SD = 17.94) and control condition (M = 67.22, SD = 14.45), t (142) = 1.23, p =.219. Just like in Study 4.1, there was a significant interaction of error type and condition, in that the difference between intra- and inter-category errors was significantly lower in the common threat condition (M = -1.87, SD = 6.53) than in the neutral control condition (M = 2.72, SD = 7.83), F(1,142) = 14.69, p < .001, η_p^2 = .09. Model fit was appropriate in both conditions (Islamist Threat: $T_1^{observed}$ = 0. 156, $T_1^{predicted}$ = 0. 100, p = .20, $T_2^{observed}$ = 29.17, $T_2^{predicted}$ = 18.24, p = .15; Racist Threat: $T_1^{observed}$ = 0. 079, $T_1^{predicted}$ = 0.104, p = .65, $T_2^{observed}$ = 16.33, $T_2^{predicted}$ = 18.75, p = .60). As hypothesized, there was a significant reduction in category memory in the Islamist (common) threat condition (Δd = .53 with the 95% credibility interval [.32, .74], p_B = <.001, see Fig. 4.2). There were no significant differences between any other parameter estimates between conditions (Table 4.2).

As predicted, there was an exclusive effect of the common threat context on decategorizing black and white US Americans relative to the intergroup threat condition. Thus, the effect was indeed driven by the content of the common threat context, not peripheral stimulus attributes. One aspect, however, deserves comment. Although Study 4.2 successfully ruled out that superficial statement attributes drive the decategorization effect, it introduced ambiguity on another front: While the results were clearly in line with our predictions, the interaction may results from two simultaneous processes: While a common enemy might indeed attenuate categorization (as in Study 4.1), it is conceivable that the common conflict did not just constitute a neutral control condition but strengthened the tendency to categorize the speaker along racial lines.

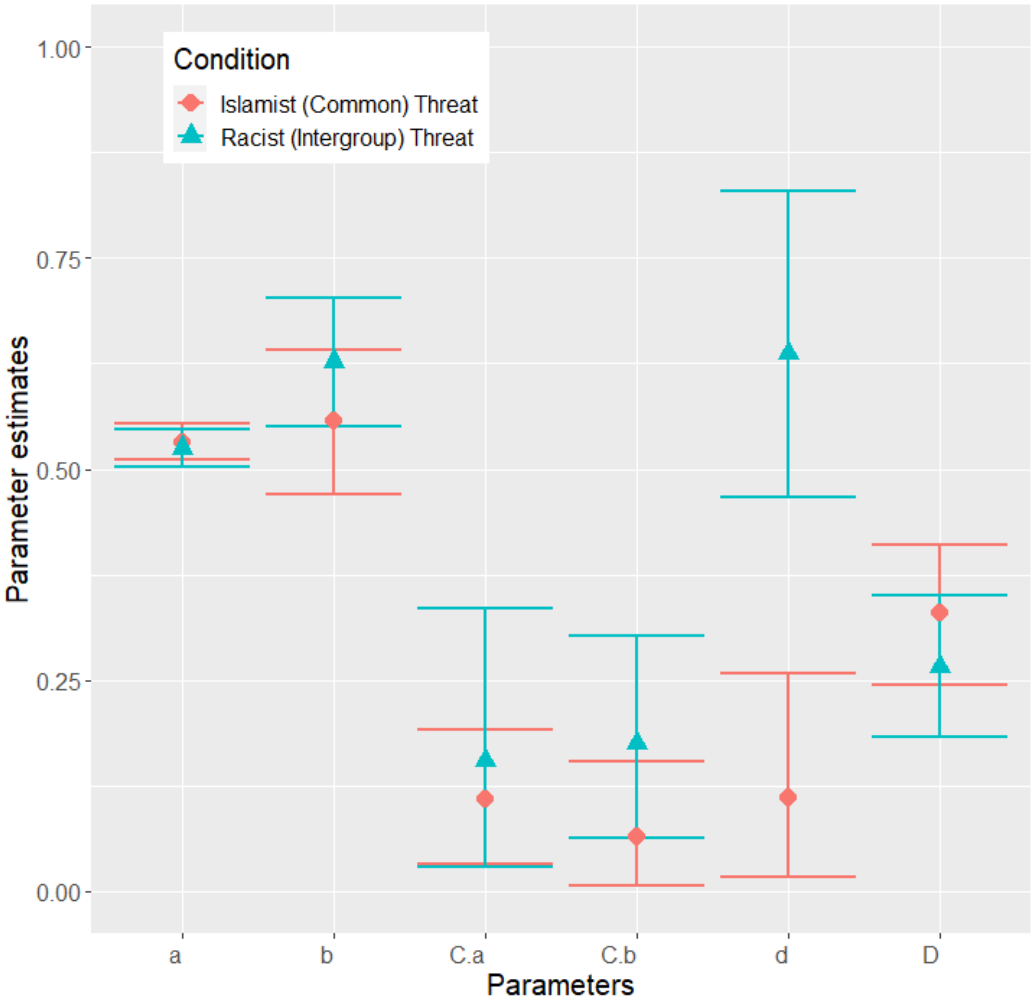

Figure 4.2. Mean parameter estimates and 95% CIs by condition in Study 4.2.

Table 4.2
Parameter Estimates and 95% CIs in Study 4.2

	Islamist (Common) Threat		Racist (Intergroup) Threat		
Parameter	*M*	95% CI	*M*	95% CI	p_B
a	.53	[.51, .55]	.53	[.50, .55]	.67
b	.56	[.47, .64]	.63	[.55, .70]	.11
C_a	.11	[.03, .19]	.15	[.03, .34]	.33
C_b	.06	[.01, .15]	.17	[.06, .30]	.07
d	.11	[.02, .26]	.64	[.47, .83]	**<.001***
D	.33	[.24, .41]	.27	[.18, .35]	.85

Note. Mean parameter estimates and 95% Credible Intervals in Study 4.2. Applied restrictions within conditions: $d_a=d_b$, $D_a=D_b=D_n$. Bayesian *p*-value (p_B) for difference in parameter estimates between conditions.

4.1.5 Study 4.3

To be able to estimate the separate contribution of a common enemy (to de-categorization) and common conflict (to categorization), we introduced a neutral condition again in Study 4.3. Thus, we constructed a neutral statement set that was matched as closely as possible to the common threat statement set in syntactic structure and semantic complexity but had an entirely different content. Additionally, to promote the generalizability of the decategorization effect, gender was chosen as target category in this study. As we kept Islamist terrorism as common threat context, target and threat categories were not associated with the same basic category dimension, race, anymore. Furthermore, we conducted the study in Germany, to be able to generalize across cultural contexts. Moreover, unlike in Studies 4.1 and 4.2 where we only had a negligible proportion of black US Americans in the dataset, we did have a substantial proportion of participants matching both speaker categories. This allowed us a first tentative look at intergroup biases. The study was preregistered at https://osf.io/hku97/register/5771ca429ad5a1020de2872e.

Method

Participants. One hundred and two participants completed the study in the lab or under lab conditions at the University of Mainz, Germany. n_{threat}= 1 was excluded because he/she did not see their data fit for analysis. Of the 101 participants in the final dataset, 26 were male, 75 female, M_{age} = 24.64, SD_{age} = 5.62, n_{threat} = 50, $n_{neutral}$ = 51. Participants received study credit for their participation. The classical error-difference measure had 80% power to detect an effect size of η_p^2 = .04 in the present study.

Stimuli/ Manipulation. Ten male (e.g. “Alexander”, “Daniel”, “Jan”) and 10 female names (e.g. “Andrea”, “Carolin”, “Luisa”) , which were unambiguous regarding their origin (German) and gender were chosen from a pre-rated set (Bonefeld & Dickhäuser, 2018).

Procedure. Same as Study 4.1. As mentioned above, we constructed a neutral statement set that was matched as closely as possible to the common threat statement set in syntactic structure but had an entirely different content. Where deemed necessary, the threat set was adapted, too, to meet the criteria and fit into the German context of the study. Examples for statements paired by syntax across conditions include: “Terrorist acts threaten Europe.”, “Urbanization threatens nature reserves.”; “Islamist organizations are loosely structured.”, “German Bachelor programs are structured by modules.”; “The international community has to negotiate with Islamists.”, “Underpaid workers have to negotiate with their employers.”

Results and Discussion

Perceived threat differed between conditions, in that participants felt more threatened in the common threat (M = 64.56, SD = 14.11) than in the neutral (M = 42.35, SD = 18.18) condition (t (99) = 6.85, p <.001, d = 1.36). Confirming our predictions, there was a significant interaction of error type and condition, in that the difference between intra- and inter-category errors, signifying categorization, was significantly lower in the common threat condition (M = 2.24, SD = 6.29) than in the neutral control condition (M = 9.65, SD = 6.31), $F(1,99) = 34.92$, $p < .001$, $\eta_p^2 = .26$.

Model fit was appropriate in both conditions (Common Threat: $T_1^{observed} = 0.143$, $T_1^{predicted} = 0.152$, $p = .53$, $T_2^{observed} = 18.09$, $T_2^{predicted} = 19.69$, $p = .60$; Neutral: $T_1^{observed} = 0.123$, $T_1^{predicted} = 0.145$, $p = .59$, $T_2^{observed} = 19.77$, $T_2^{predicted} = 18.55$, $p = .46$).

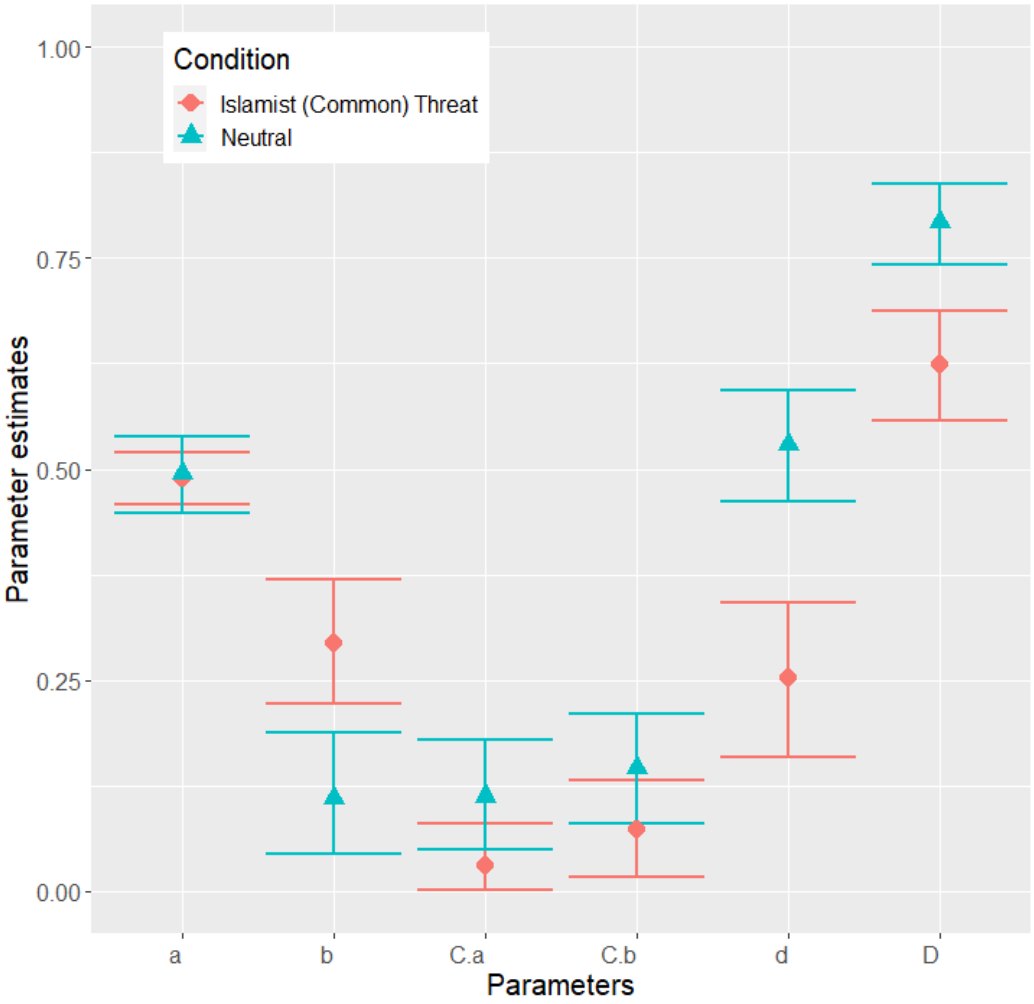

Figure 4.3. Parameter estimates and 95% CIs by condition in Study 4.3.

Table 4.3

Parameter Estimates and 95% CIs in Study 4.3

	Common Threat		Neutral		
Parameter	*M*	95% CI	*M*	95% CI	p_B
a	.49	[.46, .52]	.50	[.49, .54]	.41
b	.29	[.22, .37]	.11	[.04, .19]	.99*[1]
C_a	.03	[.00, .08]	.11	[.05, .18]	.02*
C_b	.07	[.02, .13]	.15	[.08, .21]	.05
d	.25	[.16, .34]	.53	[.46, .59]	**<.001***
D	.63	[.56, .69]	.79	[.74, .84]	<.001*

Note. Mean parameter estimates and 95% Credible Intervals in Study 4.3. Applied restrictions within conditions: d_a=d_b, D_a=D_b=D_n. Bayesian *p*-value (p_B) for difference in parameter estimates between conditions.[1]Difference is significant in the opposite direction: Common Threat > Neutral.

As predicted, the decategorization effect replicated (Δd = .28 with the 95% credibility interval [.17, .39], p_B = <.001, Fig. 4.3). Contrary to Studies 4.1 and 4.2, there was an effect of manipulation on statement detection, in that statements in the neutral control condition were discriminated more accurately (ΔD = .17 with the 95% credibility interval [.09, .25], p_B = <.001). Also, there was a difference in old/new statement memory between conditions, in that the probability that a statement was guessed to be old if there was no item memory for the statement was larger in the threat condition (Δb = .19 with the 95% credibility interval [.09, .29], p_B = <.001, Table 4.3). These two additional effects on item discrimination and old/new guessing between conditions might be attributable to a specific feature of the neutral statement set. Heavy constraints were imposed on this set, as syntactic structure and likelihood of speaker category-statement pairing were balanced, while content was supposed to be neutral regarding both categories and threat. Thus, the content of the individual neutral statements was quite diverse – much more diverse than in the threat condition. It might be that this led to increased individual statement (recognition) memory. Therefore, it is easily conceivable that statement discrimination was increased (parameter D) for these more dissimilar neutral statements. Also, due to the distinctness of the statements, participants might have overestimated their ability to recognize old statements and took their own non-recognition of a statement as a cue to guess that the statement was new (parameter b) in the neutral condition. As can be seen from the original multinomial processing tree, in principle, higher item discrimination (parameter D) might increase the likelihood of detecting categorization, as a larger proportion of answers enters this “branch” of the multinomial processing tree. However, although there logically must be a small interdependency between MPT parameters, they are usually considered largely independently interpretable, so it is unlikely that item recognition drives the decategorization effect in an essential way.

4.1.6 Study 4.4

In the previous studies, we established the decategorization effect by common threat at the early stage of spontaneous social categorization. We ruled out a range of alternative explanations and generalized the effect across target category dimensions and Western cultural contexts. Although the results are consistent and suggestive of the general principle delineated in the introduction, the chosen threat was admittedly somewhat abstract and the chosen categories potentially not as meaning-laden as that of real conflicting parties. Decategorization has often been discussed as a tool to ameliorate intergroup conflict. Although, there undoubtedly exist tremendous disparities between black and white Americans as well as German men and women, they do not constitute typical categories in conflict. To put our reasoning to a maximally conservative test, we moved to a real-life intractable and violent conflict: the Palestinian-Israeli conflict (Kelman, 1999; Rouhana & Bar-Tal, 1998). In a study conducted at the lab at Tel Aviv University, Israel we chose Israeli Jews (Hebrew speakers) and Israeli Arabs (Arabic speakers) as feasible target speaker categories. As noted above, usually, threat is studied as a cause of division between social groups. A "Common Enemy", however, constitutes the special case of a threat having the opposite effect: bringing groups closer together. As Sherif (1958) noted already early on, the uniting threat posed by a common enemy can be quite strong - however, instead of easing categorization altogether, it might simply move the separating line to a new set of categories. Therefore, he preferred the "Common Goal", which also seemed to bring groups closer together, but lacked the threatening, fear arousing component. In this study, we aimed to test all these ideas against each other. Thus, we designed 4 between-subject conditions:

1. Conflict: Context of the Israeli-Palestinian Conflict, constituting an intergroup threat
2. Neutral (Baseline): Context of everyday habits and internet usage

3. Common Goal: Context of Student life and academic goals, uniting and non-threatening
4. Common Enemy: Context of fear of illnesses & disease, uniting threat

This allowed us to test (1) whether we could also increase categorization by means of a threatening context relative to a neutral baseline or whether threat always leads to decategorization in the WSW paradigm (e.g. by means of attentional focus), (2) whether a common goal could have a decategorizing effect relative to the baseline, and (3) whether a common threat leads to decategorization over and above a common goal context. Hypotheses were (a) strongest categorization in the conflict condition, and a stepwise reduction in categorization strength across subsequent conditions, with categorization being weakest in the common enemy condition, resulting in a significant linear contrast, (b) no parallel reduction in individual person memory *C* or meaningful differences in any of the other parameters, constituting an exclusive effect of manipulation on categorization strength, and (c) the effect should occur in both Hebrew and Arab subsamples.

The respective contexts were introduced through entirely new statement sets that again varied as a function of the condition. The sets were designed from scratch to reflect the real-world situation in Israel and take into account various limitations imposed by this situation. The statements were approximately equally complex and not indicative of speaker category (like e.g. “We Arabs” would be indicative of the Arab category), so that they could be randomly assigned to speakers. There were two subsets of participants: A Hebrew speaking sample consisting of Israeli Jews, and an Arab speaking sample of Israeli Arabs (i.e. Palestinian citizens of Israel). Participants were presented with the survey in their respective native languages (Hebrew or Arabic). Having an about equally distributed sample of both Israeli Jews and Arabs provided us with the unique opportunity of investigating the decategorization effect for both groups separately. Moreover, this way, participant categories

could be matched with target "speaker" categories, allowing us to explore intergroup bias in spontaneous categorization.

As a safeguard to have enough power, we had originally included a second WSW task after the first one (which introduced another condition). This could have allowed us to double the number of participants in all experimental cells, but would have also introduced data dependencies. We have conducted all analyses reported below separately for only each participants' first WSW task, as well as with all data (collapsed across first and second WSW task). To avoid any issues with data dependency, below we present only the results for the first WSW task. The collapsed analyses are reported in the online supplementary material and yield no different results. Also, for exploratory reasons, perceived threat of context, political ideology and religiousness were assessed after the task.

Method

Participants. Israeli Jews and Arabs were recruited on campus and social networks by Israeli Jewish and Arabic research assistants respectively, resulting in a mainly student sample. Participants took part in the lab, the library or any other quiet place on desktop or laptop computers under the supervision of lab assistants. Ninety-six Israeli Jews and $N = 89$ Israeli Arabs completed the study in a lab at Tel Aviv University. Based on our standard exclusion criteria $n = 9$ ($n_{CE} = 3$, $n_{CG} = 4$, $n_{CO} = 2$), and due to technical problems, $n = 8$ ($n_{CE} = 2$, $n_{NE} = 1$, $n_{CO} = 5$) were dropped from the analysis (including these eight datasets in the analysis did not alter the results). The remaining sample consisted of 86 Israeli Jews (male/female/other: $N = 29/56/1$, $M_{age} = 25.38$, $SD_{age} = 3.27$) and 82 Israeli Arabs (male/female/other/missing: $N = 24/56/1/1$, $M_{age} = 21.65$, $SD_{age} = 1.78$). We randomly split each sample between the four conditions respectively, so cell counts ranged between 19 and 24 for each condition and language group (across language groups: $n_{CE} = 43$, $n_{CG} = 41$, $n_{NE} = 45$, $n_{CO} = 39$). Participants received 35 Shekel for their participation. The classical error-

difference measure had 80% power to detect a significant decategorization effect of the common threat relative to the neutral baseline at a minimum effect size of $\eta_p^2 = .08$ in the present study.

Stimuli/ Manipulation. Speakers were represented by male Hebrew and Arabic names, chosen for maximal distinctiveness and typicality by Israeli Jewish and Arabic research assistants. From pools of seven (Hebrew names, e.g. "Erez", "Gilad", "Noam") or 10 (Arab names, e.g. "Ahmed", "Fadi", "Hassan"), four names of each category were randomly sampled for each participant. The four new statement sets were chosen for topic and compiled and revised by the research assistants and authors according to a range of criteria. Identifying a "common threat" topic was especially challenging. The Israeli-Palestinian conflict continues to be a constant source of tension for more than 100 years (Rouhana & Bar-Tal, 1998). This intractable conflict is seen by many as one of the most enduring and pressing threats to world peace. Furthermore, it is one of the most polarizing conflicts with an almost obligatory mandate to support one side or the other, as reflected in more than 130 UN resolutions. Almost necessarily, through the lens of that conflict there is no realistic option for a third party that is threatening to Israelis and Palestinians to the same degree. Also, the superordinate category of inhabitant of the Eastern Mediterranean seemed problematic as the conflict is about each group's legitimacy to be precisely in that region. We thus decided to use a threat that is common to both groups on a level that directly threatens their only superordinate category we deemed plausible: as member of the human race, threatened by a disease: cancer.

The common threat set thus included statements such as "It is not easy to cure cancer." By doing so, we not only created a realistic common threat for the current context but also the opportunity to further bolster the generalizability of our effect. As common goal, student life and academic goals was chosen to stay close to the lifeworld of our student participants, but

also, as only a non-political topic could offer an uncontroversial common goal for Israeli Jews and Arabs. While we attempted to frame the common goal in the classic way, namely transporting a sense of "we need to help each other to achieve this" e.g. by making reference to study groups, we could not construct a whole statement set under this restriction. Therefore, for the rest of the set, we relied on "commonality", describing experiences that Israeli Jewish and Arab students would likely share. The set included statements such as "I need a quiet atmosphere to study." For the neutral condition, we also had to abolish lifeworld and working environment as topic, because living conditions differ heavily between the two groups. Instead, we chose common everyday habits and social media use (e.g. "I love my comfortable couch."). We also introduced a conflict condition, which introduced an intergroup threat similar to the one in Study 4.2. Here, we took even more care that likelihood of speaker-statement pairing was equal between categories by references to the own group (e.g. "Jerusalem belongs to my group.").

Procedure. Same as Study 4.1. Participants completed two subsequent WSW tasks. Additionally, they completed a measure of political ideology ("How would you describe your political attitude? I am… [slider scale: conservative – liberal]"), perceived threat of context ("How threatening were the statements in the task you just completed? [slider scale: not threatening at all – very threatening]") and religiosity ("How would you describe your religious feelings? [slider scale: not religious at all – very religious]"). Task completion took approx. 40 min for both WSW tasks/ parts.

Results and Discussion

Regarding the manipulation check, we had two high-threat conditions (Common Threat and Conflict) and two low-threat conditions (Common Goal and Neutral). Additionally, two of these conditions were designed to decrease categorization (Common Threat and Common Goal), whereas two were not. As intended, the two threat conditions (Common Threat: M =

31.76, *SD* = 14.77, Conflict: *M* = 36.54, *SD* = 16.54) were perceived as more threatening than the two non-threatening conditions (Common Goal: *M* = 21.19, *SD* = 21.17, Neutral: *M* = 19.19, *SD* = 10.90), *F* (1, 102) = 19.60, $p < .001$, $\eta_p^2 = .16$, while there was no significant difference in perceived threat within the high- and low-threat conditions. Regarding the binary comparison relevant to the main decategorization effect, perceived threat differed between conditions, in that participants felt more threatened in the common threat than in the neutral condition (*t* (53) = 3.52, *p* =.001, *d* = 0.95). With respect to our main hypothesis, we predicted a stepwise decategorization from the conflict condition, through the neutral and common goal conditions to the common threat condition, tested via linear contrast. As can be seen in Figure 4.4, there was a progressing decategorization effect across all four conditions (Conflict: *M* = 4.39, *SD* = 2.49; Neutral: *M* = 2.20, *SD* = 3.10; Common Goal: *M* = 2.19, *SD* = 2.84; Common Threat: *M* = 0.72, *SD* = 2.42). It was observable in both Hebrew and Arab subsamples, as well as for both (stimulus) categories separately. The linear contrast became significant in both Hebrew ($F(3,83)=10.80$, $p <.001$, $\eta_p^2=.28$) and Arabic ($F(3,78)=5.36$, $p = .002$, $\eta_p^2=.17$) subsamples, indicating that there was progressing decategorization from the conflict condition, through the neutral and common goal conditions, to the common threat condition, and confirming Hypotheses 1 and 3. In line with the second hypothesis, the manipulation had only weak and inconsistent effects on strength of individual person memory (parameter C). For the full list of parameter estimates, see Table 4.4. For the pairwise planned contrasts, (1) the intergroup threat context led to higher categorization compared to the neutral context ($T(165) = 3.54$, $p = .001$), (2) the common goal context did not show a decategorization effect relative to the neutral baseline ($T(165) = 0.37$, $p = .72$), and (3) categorization was significantly lower in the common threat relative to the common goal condition ($T(165) = 2.37$, $p = .019$).

Model fit was appropriate in all four conditions (Conflict: $T_1^{observed} = 0.258$, $T_1^{predicted} = 0.193$, $p = .28$, $T_2^{observed} = 44.93$, $T_2^{predicted} = 29.52$, $p = .15$ / Neutral: $T_1^{observed} = 0.156$, $T_1^{predicted} = 0.161$, $p = .51$, $T_2^{observed} = 24.11$, $T_2^{predicted} = 25.57$, $p = .54$ / Common Goal: $T_1^{observed} = 0.214$, $T_1^{predicted} = 0.188$, $p = .40$, $T_2^{observed} = 27.98$, $T_2^{predicted} = 26.52$, $p = .44$ / Common Enemy: $T_1^{observed} = 0.137$, $T_1^{predicted} = 0.181$, $p = .66$, $T_2^{observed} = 23.48$, $T_2^{predicted} = 23.86$, $p = .51$). Up to date, only pairwise comparisons are possible for MPT model parameters. Therefore, the data is interpreted by integrating results from both classical and MPT analyses.

As predicted, the basic decategorization effect replicated ($\Delta d = .23$ with the 95% credibility interval [.04, .42], $p_B = .01$, Fig. 4.4). Thus, the decategorization effect also showed in the context of an intractable realistic intergroup conflict. Did the common goal context lead to lower categorization than the neutral baseline context? As reported above, the error-difference measure returns a nonsignificant result. For the direct comparison of MPT parameters, the result is also nonsignificant. We thus have no evidence that a common goal context triggered decategorization, which would be interesting in itself as it does not align with Sherif's reasoning that common threat and common goals should both be very potent in attenuating intergroup boundaries. In the error-difference measure, however, noise might have rendered the effect undetectable, and in both analyses, power might have been too low to detect the effect. Above and beyond the basic decategorization effect, common threat even significantly reduced categorization relative to the common goal context ($\Delta d = .17$ with the 95% credibility interval [.01, .30], $p_B = .01$, confirmed by the error-difference measure). To the attention of the authors, these two strategies have never been tested directly against each other before. As both a common goal and a common threat aim to bring two social groups closer together, this gives rise to the idea that threat might have an additive or multiplicative effect to that end.

CIs and Bayesian ps for analysis across participant categories in Study 4.4

Neutral		Common Goal		Common Threat						
95% CI	*M*	95% CI	*M*	95% CI	p_B *(CO/CT)*	p_B *(CO/NE)*	p_B *(NE/CG)*	p_B *(CG/CT)*	p_B *(CO/CG)*	p_B *(NE/CT)*
[.47, .54]	.50	[.46, .54]	.52	[.49, .54]	.49	.34	.44	.74	.68	.28
[.15, .34]	.33	[.25, .40]	.35	[.28, .41]	.90	.32	.91	.66	.96*[1]	.80
[.14, .38]	.12	[.04, .20]	.12	[.04, .20]	.67	.98*[1]	.03*	.54	.04*	.63
[.04, .27]	.11	[.02, .22]	.11	[.04, .18]	.88	.90	.34	.54	.34	.81
[.16, .48]	.26	[.14, .35]	.09	[.01, .19]	<.001*	**.01***	**.27**	**.02***	**.01***	**<.001***
[.43, .63]	.53	[.46, .60]	.45	[.40, .50]	.24	.81	.53	.04*	.07	.85

ates, 95% CIs and Bayesian *p*s for selected comparisons in Study 4.4. X_a subset implies Hebrew speaker category, X_b ry. *significant at pB = .005. [1]Difference is significant in the opposite direction. Applied restrictions within conditions: *p*-values (p_B) for difference in parameter estimates between conditions. $p_{B(CO/NE)}$: Comparison between Conflict and : Comparison between Neutral and Common Goal conditions, $p_{B(CG/CT)}$: Comparison between Common Goal and $p_{B(CO/CG)}$: Comparison between Conflict and Common Goal conditions, $p_{B(NE/CT)}$: Comparison between Neutral and See online supplementary material for analysis of full dataset.

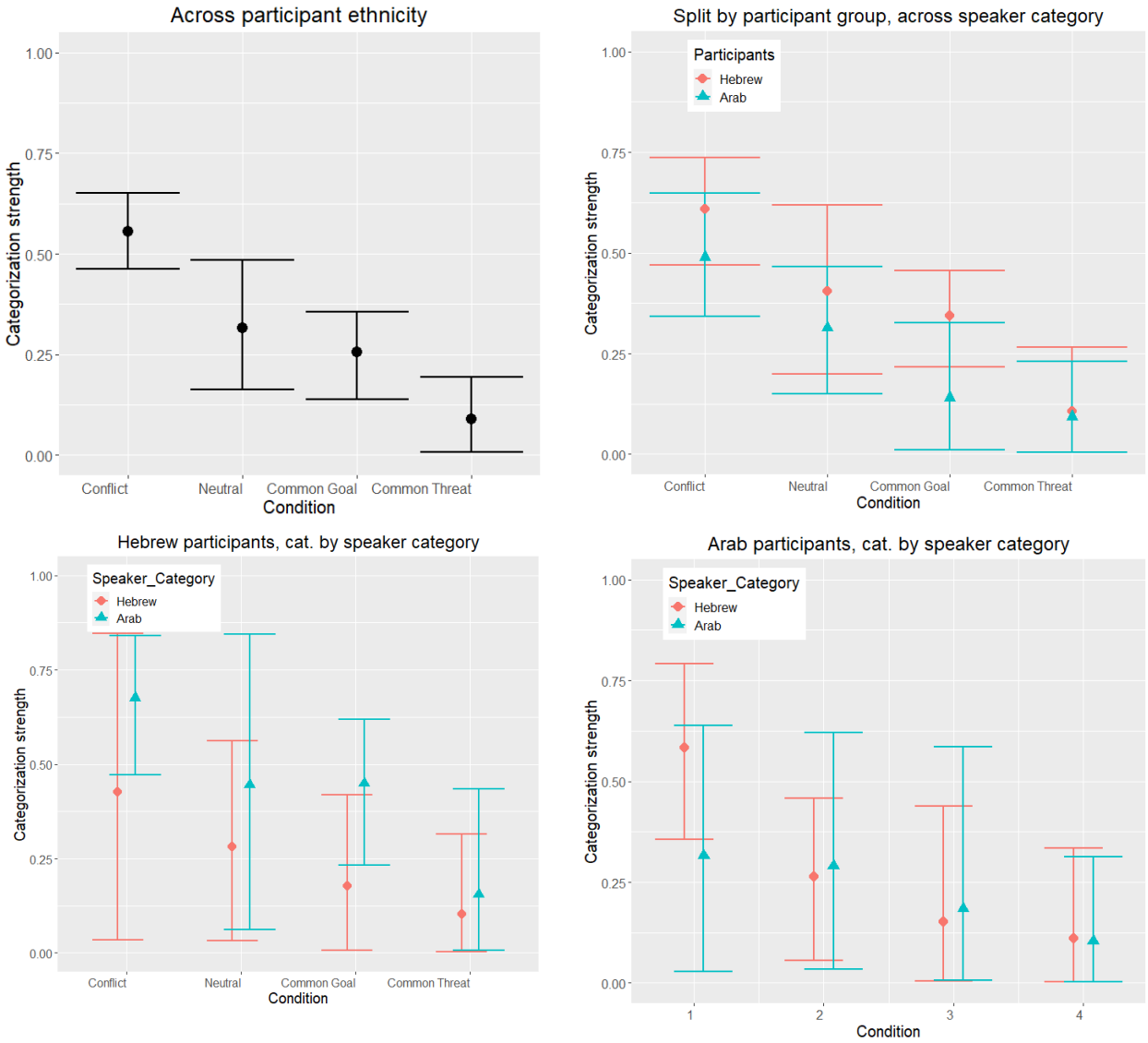

Figure 4.4. Categorization strength (Parameter d), mean and 95% CIs by condition in Study 4.4. Top left: Categorization strength by condition across both participant categories (Israeli Jewish vs. Israeli Arab participants). Top right: Split by participant category (Israeli Jewish vs. Israeli Arab participants) – effect is visible for both participant subgroups. Bottom left: Data from Hebrew participants, separate computation for categorization strength by speaker category. Bottom right: Same for Arab participants.

Alternatively, we might not have succeeded in representing the construct of Sherif's common goal sufficiently in the statement set. For him, the central defining feature was that group members had to *work together* and help each other in pursuit of the common goal. A common enemy can be evoked as a mere (shared) cognition, while it is hard to construct such a common goal in a statement set randomly ordered and assigned to speakers, and,

importantly, without referring to a "common antagonist" of any kind, that could be mistaken as a common threat/enemy. If the common goal's driving force is actually social cohesion through mutual dependency in the sense of Kurt Lewin's field theory (Bruhn, 2009; Lewin, 1943), modelling this in the WSW paradigm would pose a considerable challenge.

Additionally, we found intergroup threat increased categorization relative to the neutral baseline (Δd = .24 with the 95% credibility interval [.05, .42], p_B = .01, confirmed by the error-difference measure). This supports findings by previous studies indicating that intergroup threat increases categorization between two categories. Thus, it is not the case that any kind of threat decreases categorization in the WSW paradigm – the data suggest that it needs to be a common threat.

4.1.7 Meta-Analysis

As mentioned above, a total of 11 studies were conducted in this research program. The 7 Studies not reported here (S1-S7) are described in detail in the online supplementary material (Table 4.5 for an overview). To ensure the validity of this internal meta-analysis, this set of studies comprises all studies we conducted to test the presented hypothesis, some of the studies were preregistered and all raw datasets were made publicly available (Vosgerau, Simonsohn, Nelson, & Simmons, 2018). To determine whether the decategorization effect is robust and generalizable beyond our findings, we conducted a mini meta-analysis (Goh, Hall, & Rosenthal, 2016) by means of the R package "metafor" (Viechtbauer, 2010). The established effect size for MPT models is Cohen's ω. However, it is not available yet for the hierarchical Bayesian estimation used in this paper. Therefore, we used Cohen's d as ES metric derived from t-tests between conditions on the difference score between within- and between-category error frequencies. The random-effects model indicated a significant small-to-medium decategorization effect (d = 0.38, 95% - CI [0.16 - 0.60], z = -3.97, p < .001).

As can be seen from Figure 4.5 and the considerable heterogeneity, $Q(11) = 53.52$, $p < .001$; $I^2 = 81.52\%$, there were marked differences in the extent to which the data supported our hypotheses (although overall they clearly do). We do have some (admittedly post-hoc) speculation as to why some studies did not produce significant effects. First, some studies (S1, S2, S6) did not even produce significant categorization effects in the control groups that could then be reduced by a common threat (likely because we used overly complex crossed-categories designs or represented race not by speaker pictures but typical Black and White names, an arguably too weak manipulation). We also learned that an essay or video prime *before* the WSW discussion phase was likely not strong enough to influence categorization (S3, S7). Clearly, these reasons only appeared to us after the fact and we have no evidence for their validity. Importantly, however, integrating across all 11 studies, we still found a significant decategorization effect.

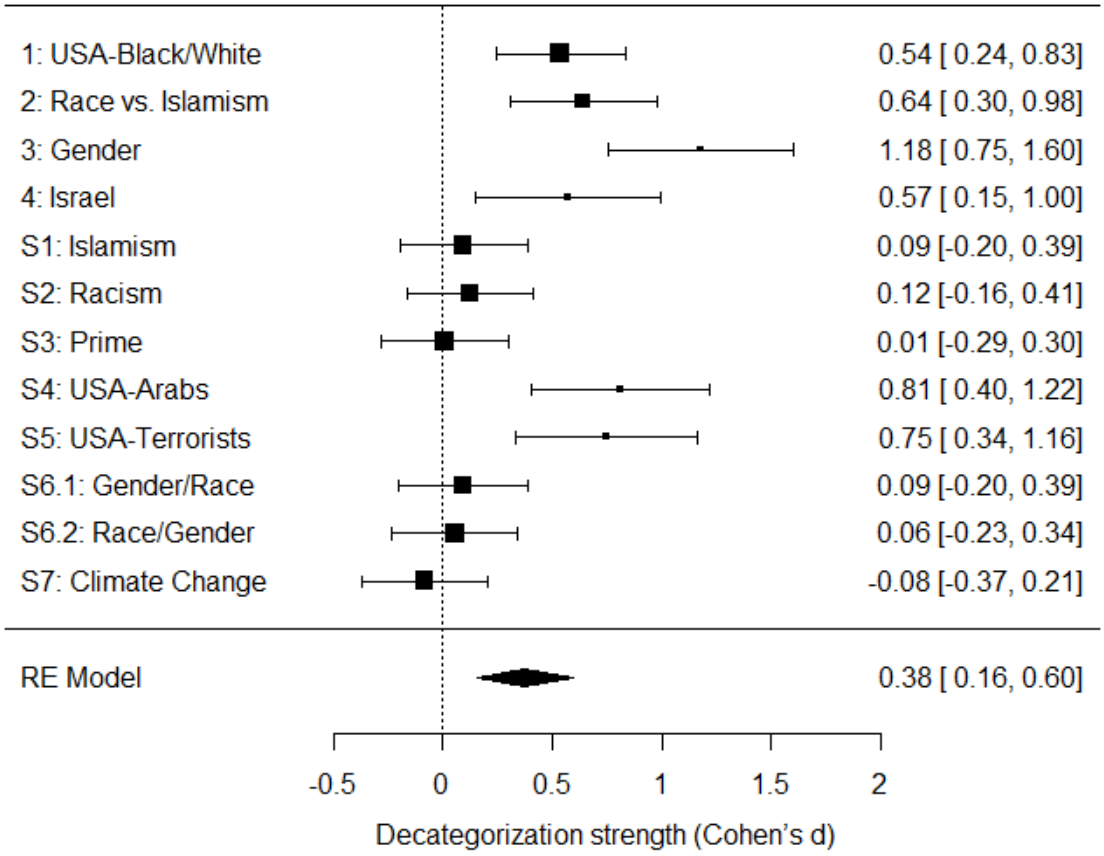

Figure 4.5. Forest plot for mini meta-analysis with random-effects model.

Table 4.5
Summary of studies in supplement.

Study	Category (speaker) stimuli	Statement topic(s)	Main hypothesis	Manipulation
S1	Black/White US names	Islamic threat vs. Race relations	Decategorization under common (Islamist) threat	Statements
S2	Black/White US names	Islamic threat vs. Racist threat	Decategorization under common (Islamist) threat	Statements (matched)
S3	Black/White US portraits	Neutral	Decategorization under threat (IS attack essay)	Essay prime (IS attacks vs. bushfires in California)
S4	White vs. Black, White vs. Arab portraits	Islamic threat vs. Neutral	Decategorization of Black, but stronger categorization of Arab category under threat	Statements, Outgroup target category
S5	White vs. Black, White vs. Terrorist portraits	Islamic threat vs. Neutral	Decategorization of Black, but stronger categorization of Terrorist category under threat	Statements, Outgroup target category
S6	Black vs. White Female, Black Male vs. Female US portraits	Gender vs. Race Threat	Decategorization under threat towards superordinate category	Target Categories and Statement Sets fully crossed
S7	Black/White US portraits	Weather related	Decategorization under threat (climate change video)	Video prime (climate change threat vs. explanation weather)

Note. Only the main decategorization effect of each study (two in study S6) was used in the meta-analysis. Additional manipulations or measures were not meta-analyzed.

4.1.8 General Discussion

Throughout a series of experiments in a variety of contexts, we have empirically established that a common threat can lead to decategorization on the early perceptual level of spontaneous social categorization. We demonstrated the effect for varying target categories (race, gender, and religion / ethnicity) and varying common threats (terrorism, disease). On a

methodological level, we varied and controlled for various stimulus properties for both speakers (textual, graphical) and statements (syntax, semantics), participant samples and ethnic groups (and therefore languages: English, German, Hebrew, and Arabic). Studies 4.2 and 4.3 were preregistered and, as power analysis for Bayesian MPT models is not yet available, conducted with sample sizes that are currently recommended for unknown effects in the field or, in Study 4.4, adhered to common practice in WSW research (Klapper et al., 2016; Klauer et al., 2014). These measures increase the trust in the generalizability and replicability of our findings. While in addition to the decategorization effect, there were also effects on item discrimination and old/new guessing between conditions in singular studies, the effect on categorization strength remains the only stable effect across studies.

The common enemy effect is a popular lay belief that is pervasive in public and political discourse. Yet, relative to the potential real-world repercussions of common threats, research on this effect has remained comparatively scarce and dispersed. We demonstrate, however, that by far not all substantial knowledge on this phenomenon has been gathered already. At the same time, this is a unique demonstration of the effect of a complex, real-world phenomenon on a deep cognitive level of psychological processes. Both the complexity of real-world common threats and the automatic nature of social categorization are reflected in the design. We will now situate the findings in the field of intergroup research on the effects of common threat and threat more generally.

4.1.8.1 Common threat, decategorization, and prejudice reduction

How do our findings relate to literature on the interplay between threat, social categorization, and prejudice reduction? The influence of threat on intergroup dynamics is usually studied in the context of intergroup threat (Riek, Mania, & Gaertner, 2006). A common finding in this line of research is that when groups threaten each other, they alienate from each other and are prone to show all sorts of negative attitudes and behavior towards

each other (Riek et al., 2006). Does this mean that the opposite is also true, in that a common threat should lead groups to display more positive attitudes and behaviors towards each other?

Generally, previous findings and rationalizations suggest that a common threat or enemy decreases intergroup bias and ameliorates intergroup behavior. Does decategorization mediate the relationship between a common enemy and lower prejudice towards outgroup members? The WSW task in a common threat context can be seen as modelling an intergroup contact setting: Two groups that have equal status within the situation engage in a conversation about a common goal (Allport, 1954; Pettigrew, 1998). Only when participants are categorized as members of their group, however, can positive contact experience be generalized to the whole group and reduce intergroup bias (Hewstone & Brown, 1986; Pettigrew, 1998). Thus, decategorization by common threat could even be counterproductive for prejudice reduction and other positive goals of intergroup contact. A possible reconciliation is proposed by Pettigrew's (1998) sequential approach to categorization in contact settings. It suggests that initial contact should occur under decategorization to reduce stereotyping, the actual contact phase should include categorization to facilitate generalization and transform into recategorization to a superordinate ingroup to achieve maximal prejudice reduction.

An alternative route to integrate decategorization by common threat into research on intergroup dynamics may be intergroup ideologies. Manipulating colorblindness vs. multiculturalism and inducing a common threat should have similar effects on spontaneous social categorization. Colorblindness should decrease categorization relative to multiculturalism (Brewer & Miller, 1984). A common threat, in turn, might lead to more colorblindness, while a common goal may lead to more multiculturalism (A. Hahn et al., 2015), especially when mutual interdependence and specialized skills are required to reach it.

4.1.8.2 Categorization by intergroup threat and construal bias

Intergroup threat increases social categorization strength (relative to a neutral baseline), as confirmed by Study 4.4. This finding is also in line with the interesting observation that the difference in categorization strength between conditions in Study 4.2, that featured an intergroup control condition, is twice as large as the difference in categorization strength between common threat and neutral context in all other three studies (quite exactly: $\Delta d_{4.2}$ = .53 vs. .21-.28), and of a similar magnitude as the difference between common enemy and conflict (i.e. intergroup threat) conditions in Study 4.4 ($\Delta d_{4.2}$ = .53 vs. $\Delta d_{4.4}$ = .47). While possible processes underlying the decategorization-by-common-threat effect are discussed in detail in Section 4.2 below, two processes underlying increased categorization under intergroup threat may be suggested here. In Studies 4.2 and 4.4, our primary goal in constructing intergroup threat statement sets was to increase the between-category difference (and, therefore, metacontrast) between the speakers. We assumed that ambiguous statements that would convey diametrically opposed attitudes depending on speaker category (e.g. “Jerusalem belongs to my group.”, in which speakers would be perceived as positioning themselves in opposing factions of an intractable conflict) would add qualitative attribute difference between speakers along category boundaries. In constructing the statements this way, however, we may have inadvertently constructed statements specifically more prone to stereotype imputation and construal bias than any statement set in other conditions. Thus, in contrast to the decategorization-by-common-threat effect, the increase in categorization we measured in intergroup threat conditions may be partly attributable to reconstructive category guessing, as explicated in Chapter 3.2.

4.1.8.3 Limitations & future directions

To advance our understanding of the conditions necessary for decategorization by a common threat, our operationalization also aimed at reducing previously implied antecedents.

Thus, we used the unobtrusive WSW measure, no candidate superordinate category was explicitly offered to the participants to facilitate recategorization, and no concerted common effort or cooperation between the two target categories was encouraged or presented to the participants. While the participants were not explicitly addressed as member of a target category, self-identification with at least one of the categories was possible as e.g. the studies featuring Black and White US American target categories also have a US American sample. Whether the uniting common threat effect also shows from a truly uninvolved observer's perspective remains open for investigation. Furthermore, our main independent and dependent variables were hardly conceptually varied across studies. While common threat was operationalized slightly differently in Study 4.4, categorization was always measured by means of the WSW paradigm. Currently, the WSW paradigm is the state-of-the-art measure of spontaneous social categorization, but this may change. To potentially transfer and replicate the effect in a different study design, it is essential to define both the constructs of common threat and social categorization with regard to the studies and trace the constructs' operationalizations within them. This will hopefully allow for the construction of bold conceptual replications and deliberate pushes to relax these scope conditions in the future.

We operationalized common threat as potentially lethal existential threat. The common threats we applied were "real", meaning that for members of the modelled categories there is a realistic chance of being (at least emotionally) affected by a terror attack or developing cancer. However, perceived likelihood (and impact) of this threat and its objective likelihood and impact may diverge in other cases. The presented studies investigate threat as a perceptual phenomenon, meaning the experience of a sense of threat (Hirschberger, Ein-Dor, Leidner, & Saguy, 2016). We expect our findings to generalize to other perceived common threats, but not necessarily to other "objective" threats (e.g. dying in a car accident objectively is a common threat to most humans but rarely perceived as such). In research on the effect of a

common enemy/ threat, the construct was often defined and operationalized differently. Early studies on common threat conceptualized threat as anticipated "loss in status" (Burnstein & McRae, 1962; Pepitone & Kleiner, 1957; Sherif & Sherif, 1953; Stouffer, 1949; Wright, 1943). This is both in line with the aggression-frustration-hypothesis and the concept of social identity threat (Branscombe, Ellemers, Spears, Doosje, & others, 1999; Tajfel, 1970). These two definitions seem distinct, but can be integrated. Defining common threat as perceived existential threat does not limit its implied effects to the physical domain. It can also include dangers to psychological "survival" in the form of identity threat, targeting culture, symbols and beliefs (Hirschberger, Ein-Dor, Leidner, & Saguy, 2016). While existential threat may be very pervasive in public references to a common threat and might also have the strongest impact, it would be interesting to see whether the effect holds for symbolic threat.

In the WSW task, categorization is defined as "grouping" (Klapper et al., 2017), i.e., categorization as simply perceiving an individual in terms of a group it belongs to rather than as an individual. This group can refer to a highly relevant category such as gender or ethnicity, but it could also be something more mundane such as "long vs. short hair" instead of female vs. male. The WSW paradigm does not tell us whether participants processed speakers by singular attributes that differed roughly by category or whether they encoded (and decategorized) "female speakers". However, the categories we used all fall under the so-called "primary" or "primitive" categories. Age, gender and ethnicity are supposedly activated and encoded automatically in all social contexts, and with equal strength (see Kurzban et al., 2001; but see Weisman et al., 2015). Therefore, these categories should be both the most salient and the least malleable of all candidate categories and thus should have provided the most conservative test for our hypothesis. To rule out a contribution of lower-level categorical encoding, we also sampled speaker stimuli randomly, and used names instead of portraits whenever feasible (see Studies S1/S2 in supplementary material on OSF for usage of name

stimuli for black/white US American speaker categories). The WSW paradigm makes quite strict assumptions regarding the input of categories: they are strictly dichotomous and often chosen to maximize the distance between target categories. Studies 4.1 and 4.2 could be criticized for that, in that the effect is not directly generalizable to portraits of people not as clearly belonging to the respective race categories. Thus, we used names instead of portraits in Studies 4.3 and 4.4. Contrary to portraits, names hardly rely on perceptual similarity but on conceptual groupability under a meaningful social target category based on learning and experience in the respective (cultural) context.

In public discourse (and scientific discourse, see Bar-Tal, Kruglanski, & Klar, 1989), narratives of the uniting force of common threat also often include references to its limited stability - once the common threat is gone, the group disperses again. In the WSW paradigm, we are only able to measure short-term effects, so the uniting force of a common goal might be less impactful immediately but prove more sustainable in the long term. A common enemy might cause (temporal) colorblindness, while a common goal activates representations of diverse resources to reach the goal, i.e. multiculturalism. Indeed, while multiculturalism seems to have larger effects on prejudice reduction in non-threatening contexts, intergroup threat reverses the pattern in favor of colorblindness (Correll, Park, & Allegra Smith, 2008; Sasaki & Vorauer, 2013). The research presented here might provide a starting point to explore these dynamics further.

As Sherif noted early on: a “common enemy” is a controversial psychological tool to make groups appear more similar. And, as social constructivism tells us, if enough people are convinced that the groups are more similar to each other, this becomes social reality. Arguably, groups united by a common enemy are inherently destructive. They can overturn regimes or discontinue a nuclear energy program and its production of nuclear waste. Or they

can support populists and demagogues in hurting arbitrary “enemies”. Thus, it is essential to investigate the dynamics underlying this effect and maybe attempt to seize leverage over it.

4.2. Why unite against? Investigating processes in the common threat effect on spontaneous decategorization between social groups

In Section 4.1, we established the decategorization-by-common-threat effect. Section 4.2 is devoted to investigating possible processes underlying that effect. By structuring and integrating the theoretical assumptions underlying research on common threat and intergroup dynamics, we identified three candidate processes and their respective theoretical premises. We then proceeded to an initial test of the central premises for each suggested process.

4.2.1 Three theoretical perspectives on common threat as unifier in social categorization

Evidence for the uniting power of common threat can be found in various studies with nearly as many different theoretical frameworks reflected in their study setups. Thus, we organized the existing literature into three theories of social decategorization under threat. Along these conceptualizations, we can distill from the literature the theoretical and empirical arguments supporting the notion that the uniting power of a common threat might already work on the early stage of social categorization that is reflected in mere shifts of similarities. All three theories feature two initially distinct target groups or categories (A) and (B) for the sake of simplicity, but all three theories can be extended to three or more target categories without making theoretical concessions. These target categories face a common threat (C), which exerts a negative force towards (A) and (B). This causes (A) and (B) to become more similar in the perception of the observer. In line with the classical definition of social categorization essentially making use of the same processes as any other categorization process (Rosch, 1978), these theories can be described without referring to the social domain at all. They describe perceptual pathways in which the introduction of an element dissimilar to two or more distinct elements increases similarity between them. We call them the redefinition theory, the reevaluation theory and the rescaling theory of social decategorization

under threat. Importantly, these three theories do not make mutually exclusive predictions regarding our main hypothesis, but all contribute to its plausibility. However, they do differ in predictions regarding boundary conditions of the effect, which will be discussed with respect to the data in the general discussion.

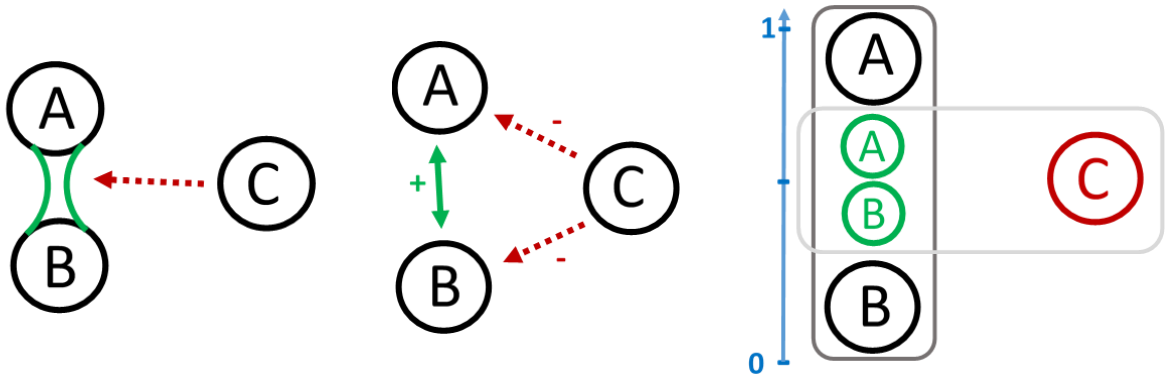

Figure 4.6. From left to right: redefinition theory, reevaluation theory and rescaling theory of social decategorization under threat

4.2.1.1 Redefinition theory

The perspective of the redefinition theory can be paraphrased by "we are not them". Here, the target object (or "matter") of the elicited perceptual change are the attributes that define categories (A) and (B) respectively. When both categories are defined by distinctive distributions of attributes, a common threat could highlight or make salient the attributes that are mutually similar or shared between (A) and (B). These new salient attributes become more strongly associated with the target category. For example, "light skin" would be a strongly salient attribute for the white American social category and "dark skin" for the black American social category. Introducing a common Islamist threat would make the otherwise only marginally relevant attribute "being threatened by Islamists" a more salient attribute for both categories, shifting their relative attribute weight and thus the respective category attribute pattern. Thus, the two attribute distributions would converge, making the two

categories and their members “more similar” in an observer’s perception. The attributes made salient this way do not necessarily have to be connected directly to the threat. For example, being confronted with an Islamist terrorist threat, the representation of black and white American categories could also shift to a stronger focus on the attribute “Christian” in opposition to “Muslim”, “peaceful” vs. “violent” etc., triggering the same process.

This is in line with the early notion that categories do not form at their boundaries, but at points of accumulation and density of attributes, making prototypes rather than boundaries the definiens of a category (Rosch, 1978). More support for this view comes from literature defining categorizing as “organizing” (Klapper et al., 2017). It defines categorical representation as the representation that has the most and strongest associations with other observed properties of a person. Several theoretical frameworks in the tradition of Social Identity Theory can be seen as implicitly assuming the redefinition process. They revolve around the idea of recategorization to a common superordinate category (Brewer & Miller, 1984; Drury et al., 2009; Feshbach & Singer, 1957; Gaertner et al., 2000; Rosch, 1978; Vezzali et al., 2016). Increasing similarity between two groups by highlighting common attributes, e.g. by means of a common threat, should ultimately cause the two groups to merge into one. In the real world, there is usually also a “candidate” superordinate category, e.g. US Americans (AB) for white (A) and black (B) Americans. According to redefinition theory and the notion of categorizing as organizing, the upward qualitative shift from (A, B) to (AB) in categorical representation would occur when the shared attributes made salient through the common threat are more strongly associated with the common superordinate category (AB) than the original categories (A, B). Implying the same process, the Common Ingroup Identity Model (CIIM, Gaertner et al., 2000) posits that intergroup biases can be reduced by recategorizing separate ingroups to an inclusive one. The Social Identity Model of Collective Resilience (SIMCR, (Drury et al., 2009; Vezzali et al., 2016) adds common threat as

candidate trigger for this process. According to SIMCR, experiencing a mass emergency as a common fate should lead to a shared social identity. This in turn is believed to activate shared goals on the cognitive level, and mutual trust as well as expecting support and agreement on a "relational" interpersonal level, leading to empowered collective action in the form of self-policing, well-wellbeing and preventing trauma, and helping behavior (Drury et al., 2009).

Redefinition theory builds on two assumptions: (1) diverse category attributes of A, B and C should be available to the perceiver at any time, and (2) the salience of these category attributes should be generally flexible and malleable by situational cues. Empirical evidence for both assumptions can be found throughout the social psychological literature. The first assumption has been the basis of influential stereotyping models such as the stereotype content model (Fiske, Cuddy, Glick, & Xu, 2002) and the 2D ABC model of stereotypes about groups (Koch et al., 2016). Regarding the second assumption, Realistic Conflict Theory and Social Identity Theory state that stereotyping reflects intergroup relations, and should thus change in a dynamic parallel to change in intergroup relations (Haslam, Turner, Oakes, McGarty, & Hayes, 1992). Early findings in favor of this idea include studies on racial stereotype change before and after Pearl Harbor (Seago, 1947) and before and during WWII (Meenes, 1943). Along the same lines, sex-role stereotypes have been found to be persistent, but also change over time (Haines, Deaux, & Lofaro, 2016; Haslam et al., 1992; Werner & LaRussa, 1985). Also, normative fit (Oakes et al., 1994) suggests a means by which stereotyping can be influenced by situational cues.

In summary, the redefinition theory is rooted deeply in social psychological intergroup research and implicitly underlies many successful lines of research and theorizing, lending strong theoretical support for our hypothesis that a common threat reduces social categorization. Yet, two more theoretical traditions are prominently represented in the

relevant literature. They can also be adapted to the present research question in the form of reevaluation theory and rescaling theory, and as such support the same hypothesis.

4.2.1.2 Reevaluation theory

At the core of the reevaluation theory of social decategorization under threat lies the idea of "The enemy of my enemy is my friend" (Kautilya, Arthasastra, Sanskrit treatise on statecraft, ca. 4th century BC), or, rephrased, "Sharing an enemy with another group makes our groups friends". It proposes that the perceived valence of the relationships (A) – (C) and (B) – (C) influences the evaluation of the relationship between A and B. Human perception is influenced (and biased) by a top-down striving for a coherent *gestalt*. According to *Balance Theory* of social perception (Heider, 1946; extended to *Structural Balance Theory*, Cartwright & Harary, 1956), there are two possible balanced states of interrelationships between three elements of a 3-element-system: Either all three relationships are positive (A company (1) produces a product (2), and both are liked by the customer (3)) or two relationships are negative and one positive (A company produces a product, but neither company nor product are liked by the customer). The third state that can be considered an interrelationship, two positive and one negative relationship, is unbalanced and unstable (A customer likes a company but not its product), holding an action potential to resolve the dissonance within the customer: The customer might attempt to find arguments to dislike the company or like the product. The same holds for the interrelationship between social groups: There is nothing startling about the US and Saudi-Arabia not liking Iran and the two states liking each other at the same time. However, there is an intuitive discord in USA having a good relationship with Qatar but not with ISIS, while Qatar has a positive relationship with ISIS (Boghardt, 2014). In Balance Theory, negative relationships were not defined to be either complement (not liking) or opposite (disliking) of a positive relationship, i.e. liking (Cartwright & Harary, 1956). While negative relationships as complements would logically lead to a non-state of

interrelationship, negative relationships as opposites would indicate a relationship indeed – which might be even stronger than an all-positive interrelationship. The common enemy effect is especially interesting in situations where an interrelationship between (A) and (B) already exists (which, in social settings, is also arguably much more frequent than group formation from scratch). This is even more so if the relationship is negative initially (e.g. in intergroup conflicts) and could become more positive by means of a common enemy. According to Structural Balance Theory, a system of three negative relationships is also defined as an imbalanced state, as the sum of the valence of its relationship is also negative (*sign of a cycle*, Cartwright & Harary, 1956). Both balance theories predict that an imbalanced state, such as a common goal of two negatively related groups or a common enemy facing two negatively related groups, would evolve into a balanced state over time. Balanced systems would remain stable. This proposition seems to be of a categorical nature: for an imbalanced state to turn into a balanced state, one or more relationships would have to switch the sign indicating their valence. Yet, a dimensional interpretation can be easily incorporated. Therefore, the reevaluation theory applies to situations in which a relationship between (A) and (B) of whichever valence is present and proposes that this relationship becomes more positive when a common threat or common goal is introduced. These notions are supported in the domain of interpersonal relationships (Aronson & Cope, 1968; Bosson, Johnson, Niederhoffer, & Swann, 2006). On the intergroup level, research on the influence of a common threat on intergroup liking (Sherif & Sherif, 1953) and prejudice reduction (Burnstein & McRae, 1962; Feshbach & Singer, 1957) could be seen in the tradition of balance theory. Reevaluation theory's explanatory power for an increase in similarity caused by a common threat is not as obvious as for the redefinition theory. It offers a process that explains an increasing positive relationship between (A) and (B) by the introduction of (C), which has negative relationships with both (A) and (B), but does not predict an increase in

similarity yet. Therefore, this model rests on the additional assumption that positivity and similarity are associated (Imhoff & Koch, 2017). Evidence for a strong positive association between valence and similarity comes from psychological fields as diverse as the mere exposure effect (Zajonc, 1968) and the density hypothesis (Unkelbach, Fiedler, Bayer, Stegmüller, & Danner, 2008). The second assumption required for reevaluation theory is that there is an intuitive preference for a good "gestalt" in social relations on intergroup level. This has not been shown empirically yet to our knowledge. However, both the striking similarity to basic phenomena in perception and the relevance of coherence in representations of the social domain (e.g. dissonance reduction, Meaning Maintenance Model, Heine, Proulx, & Vohs, 2006, or coherent neural-representational patterns as "attractors" in certain neural-network models, Freeman, Stolier, Brooks, & Stillerman, 2018) strongly suggest the plausibility of this assumption.

Structural balance theory, which lends reevaluation theory its most important concepts and propositions, is widely recognized across economics, sociology and psychology. In fact, structural balance theory not only refers to relationships between individuals or groups but also to their attributes. Along this route, one might argue that redefinition theory can also be described by structural balance theory. While a systematic theoretical integration may be highly interesting at a later stage, the present aim was to illustrate and stay true to the respective research traditions separately.

4.2.1.3 Rescaling theory

The third theory of social decategorization under threat does not focus on the salience of category attributes (redefinition theory) or the valence of relationships between the involved elements (reevaluation theory), but on the shifts in distance between (A) and (B) by the introduction of (C) in the observer's perception. Contrary to the first two theories, the rescaling theory does not build on direct interaction between the three elements A, B and C.

Astronauts often describe that from space, problems on Earth look petty - compared to the vastness of space, humans seem so alike that conflicts between them seem foolish and pointless, independent of e.g. attribute shifts or changes in interrelationship valence (Drake, 2018). Theoretically, the rescaling theory is rooted in the *metacontrast principle* (Oakes et al., 1994; Turner, Oakes, Haslam, & McGarty, 1994), according to which the perception of inter-category difference is a function of within- and between-category similarity (between each involved individual's attributes). The more similar members of a category and the more dissimilar the members between categories, the stronger the categorization. While the introduction of a common threat does not alter this metacontrast system, it temporarily alters its perception by introducing another level. Put differently, metacontrast ratio is relative similarity: the more similar members or attributes of the same category are to each other relative to the dissimilarity between categories, the higher the metacontrast. The introduction of a common threat could add a second order metacontrast: the relative similarity between the (AB) relative similarity and the (AB) – (C) relative similarity. According to rescaling theory, the much stronger meta-metacontrast ratio between (AB) and (C) should cause the metacontrast between A and B to fade in the perception of the observer. Just like the rescaling theory, the minimal group paradigm (Tajfel, 1970) evokes its effect, in this case intergroup bias, based on "mere similarity". However, even the most content-free operationalization of similarity (e.g. over-/under-estimators) uses a category attribute. The same is true for rescaling theory. This makes it very similar to the redefinition theory, which zooms in on the attribute-level micro-processes of recategorization. Contrary to redefinition theory, however, rescaling theory is based on comparison processes, but does not require assimilation (on the attribute level) as consequence of these comparison processes to explain social decategorization under threat. This requires that (A), (B) and (C) can be compared on (largely) the same attribute dimensions at all – so (A), (B) and (C) need to be social actors.

Transferred to the present research, this model would predict that the "common threat" has to be a human or social "common enemy" to trigger decategorization. In line with this notion, the original metacontrast principle is also built exclusively on social actors.

The rescaling theory assumes further that the perceptual space is (1) limited in the social domain and (2) adjusts to objects of interest that are salient in a certain situation. Support for both assumptions can be found in various branches of social psychological research, e.g. social judgment theory (Sherif & Hovland, 1961), the shifting standards model in gender research (Biernat & Manis, 1994) and the anchor effect in social comparison research (Mussweiler, 2003).

The three lines of theorizing suggest three very different pathways through which a common threat could lead to an increase of perceived similarity between its targets. The redefinition theory proposes that a common threat makes attributes salient within each category that are more similar across categories than the previously salient ones (*content similarity* proposition). The reevaluation theory proposes that a common threat introduces negative relationships with both categories, leading to a more positive relationship between them (*valence similarity* proposition). The rescaling theory proposes that a common threat introduces a meta-metacontrast which makes the metacontrast between the two original categories relatively smaller in perception (*relative similarity* proposition).

For a first approximation of the mediating processes underlying the decategorization-by-common-threat effect, we conducted an initial study testing the a-paths of the mediating processes suggested: the effect of common threat on redefinition, reevaluation and rescaling respectively.

4.2.2 Study 4.5

To approach the influence of a common threat on redefinition, reevaluation and rescaling, we designed outcome measures to the specifications of each of the three theoretical perspectives. Thus, each measure was designed to react to one of the processes exclusively. Although categorization strength was not the primary dependent variable in this study, we included a WSW task. This enabled us to implement the common threat manipulation exactly as in the preceding studies and provided us with the measure of categorization strength as manipulation check. After the WSW task, several measures were applied that were designed to capture one of the suggested processes each. As this was our first attempt, for exploratory reasons, we attempted to capture some processes with more than one measure. In these cases, hypotheses regarding only one of the measures were preregistered. Only the results of these measures are reported here. In addition to the three processes suggested above, we included tasks to measure recategorization. While we treat recategorization as a special case of redefinition in our theoretical framework, it is a process widely referenced in the literature as both major process in social categorization dynamics (Gaertner et al., 2000) and an outcome of common threat (Drury et al., 2009; Vezzali et al., 2016). All measures were implemented in a single first study, in order to identify measures that might successfully capture one of these processes and to find possible relations between them.

We report all measures, manipulations, and exclusions in this study. Final sample size was determined before data collection. Upon completion, no further data was collected. All materials, data and supplemental analyses are available on our OSF project site (https://osf.io/d3jny). This study was preregistered at http://aspredicted.org/blind.php?x=5wv6xw.

Method

Participants. Data collected by students for a course assignment resulted in $N = 171$ full datasets of Germans who took part in the study on laptops under lab conditions. All datasets ($n_{threat} = 91$, $n_{neutral} = 80$, 71 men, 98 women, 2 missing, M_{age}=31.60, SD_{age}=14.96) were included in the analysis. The post-hoc sensitivity analysis for achieved power in the MANOVA for global effects for the four main dependent variables of interest indicated 80% power to detect an effect size of $f^2(V) = .07$ in the present study. For simple comparisons between conditions, the sensitivity analysis indicated 80% power to detect an effect size of $d = .43$.

Stimuli/ Manipulation. Portraits of white and black Americans were the same as in Study 3.2. Portraits of terrorists were the same as in in supplementary Study S5 (Flade et al., 2019). To implement the manipulation in a between-design, we designed two sets of statements: The neutral statement set used first in supplementary Study S5 (Flade et al., 2019) was translated to German and used in the neutral condition (e.g., “I enjoy reading books.”, “I can run a marathon”). The statement set in the common threat condition (translated to German from the respective set in Study 3.1) introduced Islamist threat to the US (e.g. “Every Islamist can find bomb recipes online.”, “Thankfully, many Islamist organizations hate each other.”, “Drones do not defend America against the Taliban.”). See online supplementary material for complete statement sets.

Process measures. For each of the three suggested processes and recategorization, we designed one or more measures to capture them.

Redefinition. Participants were shown all speaker portraits side by side, grouped by category. Then, they were asked to name five attributes they would use to describe the respective group. If decategorization by common threat is based on the assimilation of salient attribute content, the named attributes should be more similar in the threat condition. To

quantify these between-category similarities in the open responses in the redefinition measure, all attribute lists per participants in this study were rated for similarity by $N = 31$ new participants (9 men, 22 women, $M_{age} = 29.45$, $SD_{age} = 14.04$) online.

Reevaluation. In the first measure aimed at close phenomenological approximation of reevaluation, participants were presented with one of two vignettes (manipulated orthogonally across threat conditions): "In the USA there are conflicts between black and white US-Americans. Imagine you are a US citizen living in a small town." or "Both black and white Americans are threatened by Islamist terrorism. Imagine you are an American citizen living in a small town." Then, in line with the predicted outcomes of Balance Theory, they had to indicate the tension they experienced in the vignette setting and evaluate the overall situation [negative – positive]. If reevaluation underlies decategorization by common threat, the threat vignette should be perceived as less tense and more positive than the intergroup threat vignette, while the intergroup threat vignette should be perceived as less tense and more positive in the common threat condition. As main measure for reevaluation, all binary “liking” similarities between speakers had to be indicated: For each pair of stimuli, participants were asked to estimate how much the speakers liked each other. In case of reevaluation, the resulting liking-metacontrast should be lower in the common threat condition.

Rescaling. Two measures were designed to capture rescaling. Firstly, participants rated one Black and one White speaker on all Big 5 personality traits. Under rescaling, personality ratings of black and white category members should become more similar in the common threat condition, but their rank order within and between categories should be stable between categories. As a main measure, they received a binary similarity measure similar to the reevaluation measure again, this time estimating the perceived similarity of each pair of speaker stimuli. In the threat condition, the measure included not only the 8 speaker stimuli, but also 4 portraits of terrorists (the publicly available portraits were of real terrorists, all

easily recognizable by stereotypical turbans, beards and rifles). As the main process behind rescaling was hypothesized to be a widening of the similarity scale caused by the introduction of very dissimilar Islamist terrorists, the measure differing between conditions by terrorist stimuli was designed to model this hypothesis explicitly.

Recategorization. To measure recategorization, a three-step funnel measure was implemented. Participants first received a modified Inclusion-of-Other-in-Self Scale (IOS, Aron et al., 1992; Schubert & Otten, 2002), on which they had to indicate how similar they perceived the two speaker groups to be. As a common confound in recategorization research is that the superordinate is prescribed in the task, we then asked participants to generate their own common superordinate group. To this aim, they were shown all speakers split by group on the same screen again, asked to imagine them as one group, and then asked to give that common group a name. Then, they had to indicate for each speaker, how much they perceived them to be part of that superordinate group. If recategorization underlies the decategorization effect, speakers should be perceived more as a part of the respective participant-chosen superordinate group in the common threat condition.

Procedure. Participants accessed the study on laptops in Inquisit 3. After giving informed consent, they were instructed that they were about to see “eight persons engaging in a dialogue”. Then, the participants were presented with successive paired presentations consisting of a speaker and a statement each. Statements were randomly assigned to speakers irrespective of category membership. The speaker was presented first and for 9 s, while the statement was displayed after a 1.5 s delay, so both stimuli were then simultaneously displayed for 7.5 s. There was no inter-trial break before the next stimulus pair was presented. After observing all 48 pairings, participants moved on to the surprise recall task. In the surprise recall task all statements from the presentation phase (48) and distractor set (48, in total 96 statements) were shown in random order, and participants were asked "Who said

that?" each time. They responded by ticking one of nine answer options, namely the eight portraits and the option "None. This statement is new."

Afterwards, participants were presented with the redefinition, reevaluation, rescaling and recategorization measures in the order described above. Participants finished the study by indicating their gender and age and filling out a short check question about the setting in which they completed the study, in which they had to describe their surroundings in a few words.

Results and Discussion

First, we report the effect of the common threat on decategorization in the WSW paradigm, a measure serving as manipulation check for the basic effect in this study. Indicating reduced categorization under threat, the difference between intra- and inter-category errors was significantly lower in the common threat condition ($M = 2.70$, $SD = 4.70$) than in the neutral control condition ($M = 7.06$, $SD = 4.70$), $t(1,181) = 6.00$, $p < .001$, $d = .93$, 95% - CI [.61-1.25].

We then computed Bayesian hierarchical latent-trait MPT models (Klauer, 2010) by means of the R package 'TreeBUGS' (version 1.2.0, Heck et al., 2018) for both conditions separately. In addition to the restrictions suggested by Klauer and Wegener (1998), category memory parameters d_a and d_b were restricted to be equal within conditions, as there were no a-priori hypotheses regarding a difference between the two (see also Klauer et al., 2014). Model fit was appropriate in both conditions (Common Threat: $T_1^{observed} = 0.087$, $T_1^{predicted} = 0.084$, $p = .49$, $T_2^{observed} = 17.76$, $T_2^{predicted} = 12.31$, $p = .21$; Neutral: $T_1^{observed} = 0.165$, $T_1^{predicted} = 0.095$, $p = .13$, $T_2^{observed} = 17.55$, $T_2^{predicted} = 13.17$, $p = .27$).

As hypothesized, there was a significant reduction in category memory in the threat condition compared to the neutral condition ($\Delta d = .31$ with the 95% credibility interval [.21, .41], $p_B < .001$). Additionally, there were significant differences in all other parameters

between conditions (see Table 4.6). Specifically, participants remembered neutral statements more than threat statements (D), remembered individual speakers who said neutral statements better than speakers of threat statements ($C_{a/b}$), guessed for more neutral statements that they had not seen them already (b), and guessed for more neutral statements that they had been said by a black speaker (a), albeit categorization strength d (and person memory for white speakers, C_a) differed most between conditions.

While the basic decategorization-by-common-threat effect replicated, the present pattern of all but the guessing parameters decreasing in the common threat condition could indicate that the threat statements lured participants' attention away from the speakers altogether, reducing both memory for individual speakers as well as for the speaker categories. Alternatively, a mixture of reasons described in Studies 4.1-4.4 may be responsible: Person memory ($C_{a/b}$) decreased in the common threat condition as in Study 4.1, statement memory (D) increased and guessing a statement to be new (b) decreased in the neutral condition as in Study 4.3. Regarding the WSW data, it may also be interesting to note that there was an ORE-like asymmetry in person memory in both conditions, in that white US American speakers were remembered individually more than black US American speakers by the (white) German participants (Common Threat: $\Delta C_{a/b} = .13$ with the 95% credibility interval [.07, .18], $p_B < .001$; Neutral: $\Delta C_{a/b} = .39$ with the 95% credibility interval [.32, .47], $p_B < .001$). Thus, this study remains the only one featured in the categorization and individuation asymmetry meta-analysis (in Study 2.1, as unpublished, 2019) that displayed an ORE-like asymmetry within a WSW task with speaker repetitions, and the only one that featured German participants in a black-white category context. It remains to be seen whether Germans are so unaccustomed to black faces that even repetition cannot compensate their (perceptual, motivational, or memory-related) outgroup bias.

Table 4.6
Parameter Estimates and 95% CIs in Study 3.5

	Common Threat		Neutral		
Parameter	*M*	95% CI	*M*	95% CI	p_B
a	.47	[.45, .49]	.43	[.39, .46]	.97*[1]
b	.40	[.34, .46]	.32	[.26, .37]	.97*[1]
C_a	.14	[.09, .19]	.48	[.41, .55]	<.001*
C_b	.01	[.00, .04]	.09	[.05, .13]	.001*
d	.13	[.07, .19]	.44	[.36, .52]	<.001*
D	.56	[.52, .61]	.64	[.60, .68]	.01*

Note. Mean parameter estimates and 95% Credible Intervals in Study 4.5. Applied restrictions within conditions: $d_a=d_b$, $D_a=D_b=D_n$. Bayesian *p*-value (p_B) for difference in parameter estimates between conditions. [1] Difference is significant in the opposite direction: common threat > neutral.

Regarding the redefinition hypothesis, characteristics attributed to speakers in the common threat condition were not perceived as more similar between categories as in the neutral condition (Common threat: $M = 4.05$, $SD = 1.32$; Neutral: $M = 3.98$, $SD = 1.18$; $t(164) = 0.36$, $p = .72$, $d = .06$. 95% CI [-.25-.36]). Regarding the reevaluation hypothesis, we computed the metacontrast ratio from all binary liking ratings. The liking-metacontrast did not decrease due to the common threat (Common threat: $M = 1.14$, $SD = .27$; Neutral: $M = 1.10$, $SD = .24$; $t(169) = 1.07$, $p = .29$, $d = .16$. 95% CI [-.15-.46]). There were also no differences between conditions on tension ratings and evaluation of the situation vignettes. Regarding the rescaling hypothesis, we computed the metacontrast ratio from all binary similarity ratings. The similarity-metacontrast did not decrease due to the common threat, but increased slightly instead (Common threat: $M = 1.32$, $SD = .51$; Neutral: $M = 1.15$, $SD = .47$; $t(169) = 2.25$, $p = .03$, $d = .33$. 95% CI [.02-.63]). Regarding the recategorization hypothesis, commonly threatened speakers were not considered to belong more to the superordinate ingroup assigned by the participants than those in the neutral condition. Descriptively, the effect pointed into the opposite direction (Common threat: $M = 5.46$, $SD = 1.09$; Neutral: $M = 5.79$, $SD = 1.17$; $t(169) = 2.25$, $p = .05$, $d = .29$. 95% CI [-.009-.60]). The preregistered MANOVA confirmed these results. Of the additional measures, only the Inclusion-of-Other-in-Self Scale displayed

a significant difference in the predicted direction. In the common threat condition, participants indicated a stronger overlap between the groups than in the neutral condition (Common threat: $M = 4.53$, $SD = 1.40$; Neutral: $M = 3.90$, $SD = 1.49$; $t(169) = 2.84$, $p = .005$, $d = .25$. 95% - CI [.05-.55]).

Thus, none of the measures indicated preliminary evidence for any of the processes suggested above. Still, it would be unwarranted to claim that there must be another process responsible for the common-threat-by-decategorization effect based on these initial results only. At this early empirical stage, there are study design issues to be considered. Also, the results need to be seen in the context of the position of this study within a research cycle.

Interpreting the results poses several challenges. While we were able to isolate an effect of common threat on decategorization selectively in Studies 3.1 – 3.4, in the present study, common threat seemed to also affect all other cognitive process parameters. There are several potential explanations for this. Firstly, to maximize the decategorization effect, we used statement sets similar to Study 4.1, in that statement sets differed also in e.g. semantic complexity. Secondly, students collected the data from various personal acquaintances that might have been unaccustomed to this kind of task and therefore might have been impacted more by the manipulation (e.g. in their attentional focus, as discussed in previous studies). Thirdly, this was the first study in the decategorization-by-common-threat research line in which participants could not naturally self-identify with one of the speaker categories, as they were both deemed US American, which may have also impacted the results. Regarding the process measures, the most apparent limitation is the long line-up of diverse tasks in a single study. Measures and participants' answers to them might have influenced subsequent measures, blurring effects. As a backrating of the open response was required for the first measure, this may also have introduced additional noise. This challenge might be met by scrambling the open responses across speaker categories to remove response dependencies

(participants may have written down “opposite” adjectives to describe the two speaker groups). Generally, it may be warranted to increase power in three ways: By increasing sample size, by investigating the potential mediating processes and their measures in separate studies, and by possibly moving the WSW assignment phase to the end of the study. Yet, participants in the common threat condition indicated that they perceived a greater overlap between the two speaker categories on the IOS measure at the very end of the study. Although this measure does not point to any of the processes specifically, and the effect would need to be replicated to become a reliable basis for interpretation, it may tentatively indicate that de-accentuation took place and can be traced beyond the time interval occupied by the WSW assignment phase – we may just need more power to detect and dissect it.

Generally, Study 4.5 is the first tentative study embedded in a larger theoretical framework - and strong theorizing is not swayed easily by a single study’s non-significant results (Fiedler, 2017). When seen in the larger frame of “loosening” and “tightening” processes within the creative cycle of scientific progress, this study is thus roughly located at the end of a loosening process of creative theorizing and at the beginning of a tightening process of statistical hypothesis testing (Fiedler, 2018) that may lead to substantial new insights into the nature of social categorization under common threat.

Chapter 5 – General Discussion

At the beginning of this research endeavor, we wanted to learn more about the nature of social categorization, namely the way it "reacts" to different influences and settings. We found that social categorization is initially asymmetrical, in that outgroups are categorized more than ingroups, but becomes symmetrical when we encounter the target persons frequently. Also, contrary to a frequently encountered lay belief, we do not seem to automatically individuate less when we categorize more. In most cases, we can individuate and categorize simultaneously, and independently of each other. Furthermore, initial data suggests that categories that we have stored in the back of our heads do not only come to mind when we find that our stereotypes are either confirmed or challenged. Categories may be self-reinforcing to the degree to which stereotypes associated with them replace information gaps and become "false memory". This does not mean, however, that social categories are impervious to external influence (other than being challenged by their own kind in the form of category competition, Klauer et al., 2014). Specifically, a common enemy might have a more profound effect on intergroup perception than increasing our sympathy for outgroup members – it makes us focus less on the categories that usually divide us. As neither concept has been studied in its relation to social categorization before the present research, all three approaches may lack a certain degree of subtlety: We manipulated symmetry by presenting category exemplars once or six times, use as many stereotypes pre-held by the participants as possible, and use existential threat instead of a common goal or attitude to reduce categorization strength. Still, such "proofs of existence" may presently still be valuable to delineate the phenomenon of social categorization. The findings reaffirm social categorization as a multifaceted phenomenon that might be, to refer to Ockham's razor, more complicated than a metacontrast based on similarity-dissimilarity (Bruner et al., 1956), but also simpler than

described in current models of impression formation (Brewer, 1988; Fiske & Neuberg, 1990), in which multiple instances of categorization are described in a single categorization process. All in all, social categorization seems to be a quite paradoxical phenomenon: Both symmetrical and asymmetrical, both self-perpetuating and malleable, a concept of "either-or" grounded in processes of "both-and", the core concept eluding evaluation, but its outcomes ranging from indispensable to morally questionable. In the following, we revisit social categorization as a phenomenon between similarity and self-identification with respect to the three research lines described in this work.

To sharpen the concept on the phenomenon of social categorization, it is central to establish it in relation to closely related concepts. So far, social categorization seems hard to separate from similarity (i.e. the metacontrast ratio, but see Thibaut, Dupont, & Anselme, 2002) to one side and intergroup categorization to the other side. This is also reflected in our studies and theoretical elaborations, as can be noted e.g. by the use of the term "intergroup categorization" instead of "social categorization" in Chapter 2. Similarly, in Chapter 4, we discuss the discrepancy between theoretically claiming that social categorization does not require self-identification on the one hand and not constructing the studies in a way that considers or tests this claim on the other hand. The challenge of dissociating social categorization from perceived similarity and intergroup categorization might also positively be a sign of their conceptual closeness and might speak to the concept's phenomenological placing. Thus, in search for a more holistic conceptualization of social categorization, it might be fruitful to take on this challenge.

5.1 Decomposing social categorization variance

Deliberations on the nature of social categorization seem to eventually always circle back to its relation to similarity and metacontrast. Both WSW and ORE face perception task claim

to measure social categorization processes and converge in producing similar patterns based on a metacontrast operationalization. This underlines the importance of the categorization-metacontrast relation, and theoretically decomposing the variance attributed to social categorization in the WSW paradigm may define it more precisely. When we consider social categorization to be a cognitive process based on perceiving similarity-dissimilarity among multiple stimuli, the WSW paradigm models social categorization well (e.g. by using spontaneous memory for an "observed discussion" setting that lacks category labels). Moreover, in contrast to real-world situations, it enables us to measure these "natural" spontaneous social categorization processes in an almost isolated fashion. Thus, by attempting to theoretically decompose the variance contained in the d-parameter value (i.e. social categorization strength), we might not just be able to critically evaluate the measurement validity of the WSW paradigm, but differentiate between conceptually distinct perceptual and information processing components that produce what we measure (and define?) as "social categorization". Four components are identified in the domains of information ecology, perception, interpretation, and evaluation, and listed according to their assumed position between bottom-up and complex top-down processing.

The most fundamental component of variance that is measured as categorization in the WSW paradigm is mere binary stimulus similarity in the information ecology. Researchers often sample WSW "speaker" stimuli that they consider to unambiguously belong to one of two categories (White / Black) along a category dimension (race; see e.g. the methodology reported in Chapter 4, and MDS analyses of binary rating data collected e.g. in the WSW studies published in Imhoff et al., 2018). This may make the metacontrast ratio especially salient or even exaggerate it, possibly leading to measured social categorization based on (mere) accurate perception of the similarity structure between stimuli. Thus, no perceptual "lumping and splitting" (Zerubavel, 1996) is required, no social category "tag" has to be

assigned top-down to the speakers - in short, no social categorization has to occur for this variance to register as "social categorization" on the d-parameter. Whether this is problematic, depends on research question. This similarity component of variance merely reduces power when attempting to manipulate social categorization by some factor in an experimental design – everything else equal, this variance should be of a similar magnitude in all conditions. However, if the research question inquires whether people categorize on a given category dimension *at all* (e.g., Imhoff et al., 2018; Klapper et al., 2016), this may lead to inaccurate conclusions (see e.g. reanalysis of Klapper et al., 2016 in Degner et al., 2020), in that the mere ("accurate") perceptual image of an ecological metacontrast may register as social categorization. This similarity-based variance, or "ecological metacontrast" cannot easily be eliminated by means of study design. Including it as a covariate, however, could reduce this problem and increase the power to detect a difference in "categorization proper" (Degner et al., 2020).

The second component is comparable to accentuation (Tajfel & Wilkes, 1963). It is still based on perceptual processing of ecological attribute distributions only. This similarity distribution, however, is superimposed with an "empty" social category. For example, a distribution of visual features of age in portraits could be accentuated to "the wrinkled" and "the un-wrinkled" similar to object categorization. Another example would be the categorization of two groups of aliens (or unknown ethnicities) that we perceive as visibly distinct. We have neither labels nor stereotypes for them (yet), but they differ in markers that have differed between ethnicities in our learning history (e.g. skin color). Thus, we know that these are two "ethnicities", and that is enough for accentuation to occur. This basic, content-free categorization should be amplified by category labels and/or stereotypes, if they are available to the categorizer. Given that categorization and stereotyping are separate processes, with categorization preceding stereotyping by a hair (Sherman et al., 2011), this form of social

categorization (“initial categorization”, Fiske & Neuberg, 1990; “identification”, Brewer, 1988) would not necessarily be preceded, though likely followed by (psychological) stereotyping. This is the process we initially argued for in the introduction to Chapter 3 - Construal Bias. If stereotypes imputed into the statements shape the perception of the speakers already in the discussion phase, and this augmented representation of the speakers is used to infer their category in the assignment phase, this would be the corresponding variance component.

The third variance component has been paraphrased as reconstructive category guessing within the WSW framework (Klauer et al., 2014; Klauer & Wegener, 1998) and requires stereotype application in the interpretation of perceived information. As described in Chapter 3, these stereotypes can be imputed into information gaps in a target person’s communication. By recalling the thereby biased statements, the category that was associated with the target person earlier on can be re-inferred. As argued earlier on, stereotype imputation may be possible for nearly all statements to different degrees (and not only stereotype-consistent ones), so this variance component may well contribute to social categorization being reinforced outside of and as measured within the WSW paradigm.

The last variance component contains all outcomes of a perceiver’s self-identification with one of the categories, that do not already fall into one of the former components. In social categorization’s special case of intergroup categorization, self-identification could provide both a category label (“me” vs. “not-me”, variance component 2) and a kind of stereotype (“just like me” vs. “not like me”, variance components 2 / 3) to enhance processes in other variance components. In addition to increasing importance and relevance of the target category dimension, intergroup contexts are closely linked to intergroup bias in the form of ingroup favoritism and outgroup derogation. This differential evaluation might be motivational (Hewstone et al., 2002) or perceptual in nature (Alves, Koch, & Unkelbach,

2018). This component could be primarily responsible for asymmetrical effects like early asymmetrical categorization (Chapter 2), and gain influence with increasing prejudice on the side of the perceiver, when evaluation correlates strongly with the category assigned to the target person (Johnson, Lick, & Carpinella, 2015).

Which of these variance components can be considered "categorization proper"? Most theoretical perspectives agree that mere unbiased perception of a similarity structure among stimuli in an ecology (component 1) is not (Pietraszewski & Schwartz, 2014). Not all attributes that differ between people, such as eye color, become meaningful categories ("social categorization is a distinctly different process than simply noticing the differences between people" Pietraszewski & Schwartz, 2014, p. 45). Strictly speaking, this means that "categorization as representing", in which any mental representation is considered a categorical one (Klapper et al., 2017), cannot be considered social categorization at all. Arguably, all other variance components are of "social categorical nature".

These three variance components all share one feature that may qualify them to be considered "categorization proper". They are all different manifestations of "pre-held belief" (Oeberst & Imhoff, 2020). This pre-held belief is imposed on an ecological similarity structure that may or may not be arranged in a metacontrast. It might even partly replace ("explain" or "interpret") variance previously occupied by an ecological metacontrast pattern. This begs the question whether people differ in the kinds of social categorical representation they hold as a pre-held belief. We all categorize, but some may do so by using "empty" categories, and others may use prejudice-infused stereotypes. The same applies to different category dimensions. It would also be interesting to know which of these four components is influenced by common threat or construal bias. Construal bias might use and influence pre-held belief (i.e., "categorization proper"). On the contrary, "mere" similarity perception might be more malleable than "empty" categorical pre-held belief, which might itself be more

malleable and adaptive to new environmental challenges than pre-held belief that is underlaid with labels, stereotypes or prejudice. Common threat may primarily influence the most malleable variance components, i.e. similarity perception and "empty" categorization. In short, there might be two major kinds of (social) categorization proper that differ in composition, complexity and malleability: pre-held belief induction and real-time accentuation.

Social cognition has had a somewhat paradoxical perspective on the concept of social categorization. On the one hand, it is necessary, ubiquitous and beyond evaluation, just like object categorization. On the other hand, it is considered the "root of all evil" (Oakes, 2008), in that it leads to stereotyping, prejudice and discrimination. Differentiating between the various variance components of social categorization might help explain this. As laid out above, mere, un-labeled, content-free accentuation based on an ecological meta-contrast in the stimulus structure, perhaps accompanied by valence-neutral stereotypes, can indeed be considered unproblematic. The potential for trouble may stem from the kind of meaning social categorization produces. To gain predictive value in social contexts, stereotypes are probably more valuable if they are psychological, if they contain ostensibly prototypical thinking styles, traits, reaction schemata, and behaviors. Those stereotypes are also more valuable if they are associated with social categories that are easily recognizable (or, at least, for which a social representation of a - preferably visual - prototype exists). If target persons object to such categories and stereotypes they are assigned to, and the (mostly negative) evaluation that is often bestowed on them based on these schemata, intergroup tensions may arise easily.

5.2 On the role of self-identification in social categorization

The most often remarked difference between social categorization and (object) categorization is that social categorization can (and often is) accompanied by self-identification. This can lead to outgroup homogeneity (Park & Rothbart, 1982) and outgroup derogation (Otten, 2016; Tajfel et al., 1971). Some work even seems to imply that social categorization is always intergroup categorization (Kawakami et al., 2017). Thus, phenomenologically, is social categorization only the interaction of (object) categorization and self-identification? Especially the asymmetry observed in categorization of previously unseen exemplars (Chapter 2) seems to suggest that self-identification plays a role very early on: outgroup members are initially categorized more than ingroup members. Yet, we cannot be sure. While we claimed to study social categorization which we defined to be independent from self-identification, we mostly did so in intergroup settings. Even in Study 3.5, although Americans comprised the speaker categories and participants were German, participants could well have self-identified with the white American over the black American category. Thus, this subject needs further investigation. One candidate for disentangling these phenomena may be the temporal asymmetry-to-symmetry dynamic of social categorization (Chapter 2). Is the asymmetry-to-symmetry effect unique to social categorization or does it generalize to the non-social domain across both paradigms? If it does generalize, the asymmetry may stem from differential perceptual expertise that applies to outgroups, but also to unfamiliar objects. If the asymmetry applies uniquely to social categorization, it might be motivated (Hugenberg et al., 2010). As ingroup faces belong to people that we expect more interaction with, we may need a more fine-grained evaluation of them as individuals, in order to specify our expectations of them (Chance & Goldstein, 1996; Hugenberg et al., 2010). While there is some evidence that the perceptual expertise account is more prominent in explaining outgroup homogeneity (Hugenberg et al., 2007; Meissner & Brigham, 2001), only the motivation

account would speak to a qualitative phenomenological difference between categorization and social categorization.

5.3 On the notion of "primitive" social categories

> „The full repertory of innate categories - a favorite topic for philosophical debate in the19th century - is a topic on which perhaps too much ink and too little empirical effort have been spilled." – Jerome Bruner (1957, p. 125)

The notion that some dimensions of categorization are more important and more deeply engrained in human cognition than others is widely accepted in social psychology and beyond (Fiske et al., 2018; Fiske & Neuberg, 1990; Hirschauer, 2014). The ones named most often are gender, ethnicity, and age (Brewer, 1988; Fiske & Neuberg, 1990; Hirschauer, 2014; Kurzban et al., 2001). Moreover, cultural studies have historically developed fully separate sciences (or fields) for each of these categorical dimensions, most prominently ethnology and Gender studies (Hirschauer, 2014). While this is not the case in social psychology (yet), it may be mirrored in the current institutionalization of social psychological gender research, and the increasing incidence of intersectionality research (Cole, 2009) and ageism studies (North & Fiske, 2012) within the field. From the perspective of cultural studies, this structural separation makes perfect sense, as categorizations by age, gender and ethnicity are culturally expressed very differently (Hirschauer, 2014). In fact, their manifestations might appear too distinct to allow for a fruitful comparison in the sense of structural alignment (Gentner & Markman, 1994), and, therefore, meaning making. While cultural studies may be more concerned with e.g. differential goals (separation between ethnicities vs. matchmaking between genders), they may find a shared interest with social psychologists in the nature of attributes or stereotypes that inform categorization (Hirschauer, 2014). Specifically, while cultural studies may over-emphasize qualitative differences between attributes and

stereotypes, social psychologists often seem to disregard or theoretically over-generalize across these differences. Social-cognitive research often lets attributes and stereotypes free to vary between (and within) participants without manipulating and measuring their content (as in Chapter 3 - Construal bias). Research that contains explicit attribute and stereotype content, on the other hand, seems to focus on the perception of visual attributes in relation to categorization (Imhoff, Woelki, Hanke, & Dotsch, 2013; Klapper et al., 2016; Yang & Dunham, 2019). Exceptions are research on the influence of accent on categorization (Pietraszewski & Schwartz, 2014; Rakić, Steffens, & Mummendey, 2011), or studies that manipulate individual psychological stereotypes (Macrae, Stangor, & Milne, 1994; Verhaeghen, Aikman, & van Gulick, 2011). While visual attributes and stereotypes are usually studied with non-semantic portrait stimuli (Miller, 1988; Yang & Dunham, 2019), attributes and stereotypes that are manipulated or studied semantically often seem to be psychological in nature, e.g. personality traits or differences in behavior associated with certain categories (Macrae et al., 1994; Verhaeghen et al., 2011). As "primitive" categories are usually easily distinguishable for contemporary humans and their culturally pre-formed cognition and world (e.g. by skin color, prominence of jawline, wrinkles), social psychology finds itself in the situation in which most studied categories are visual, and most associated stereotypes are psychological in nature. Exceptions are the treatment of trustworthiness as a category and its visual manifestation as stereotype (Klapper et al., 2016, but see Degner et al., 2020), and the ABC model of stereotype dimensions, in which categories on the level of socially represented social groups (lawyer, punk, mother) serve as sources for stereotypes (rich, leftist, common), that then inform stereotype dimensions on a meta-level (agency, belief, communion, Koch, Imhoff et al., 2016). On these stereotype dimensions, (psychological) categorization takes place again (Imhoff, Koch, Flade, 2018). This fits to the idea that initial, "basic" categorization is visual (i.e. based on visual attributes), but since

social categorization's primary goal is meaning making and prediction of other's behaviors (that would count as psychological stereotypes), categories *of* social categories should be psychological rather than visual. Adhering to visual stereotypes about the category one is assigned to (uniforms, scientist's fly, housewife's apron; *impression management*, Goffman, 1959; Tedeschi, 1981), might thus be a way to signal one's psychological attributes / stereotypes, if they seem desirable. If they are not, some choose to alternatively signal that they are exemplars of a category other than the initially salient one, e.g. by crossdressing or dressing youthfully as an elderly person. This may visually communicate that stereotypical behavior adhering to the perceivers initial categorization should not be expected from this person.

That basic category distinctions are mainly visual may wrongfully suggest that they are somehow "meant to be", i.e. rooted in evolution or the like. Historiography on the social construction of ethnicity and gender may object. For example, in medieval Europe, white skin was not considered good and black skin bad – there were three skin colors (white, red, black), each had a temper associated with it and a "good" skin tone was one mixed from at least two of them, as their owner was assumed to have a "balanced temper" (Groebner, 2003). Also, in ancient Greece, the (only) gender was a continuum of "different shades of male" (Laqueur, 2003). Thus, while there may be good reasons to use so-called "primitive" categories as exemplars in social categorization research, this may also hold the danger of a certain "naïve empiricism" (Scherr, 2020, p. 3), in that such studies may reproduce and reinforce historically imparted category dimensions instead of examining them critically (Scherr, 2020). Within the scientific discourse beyond social psychology, it may also be controversial to generalize findings obtained on "primitive" categories to all instances of social categorization. Thus, a stronger focus on conceptual replication may be key to convincing fields more skeptical of

generalizability of the amenities of thinking in terms of overarching processes beyond context and content.

5.4 Conclusion

Our perception of the real world is dimensional, but we can only understand these perceptions with the aid of categories. Social categorization is an intricate mechanism that allows us to do that. As it allows us to understand, it may also allow us to understand the nature of social categorization itself and thereby find ways to circumvent its negative side-effects. Social categorization can be indispensable and malleable at the same time. Likewise, we can both admire the intricacy of our own cognition – and be critically aware of its conclusions.

References

Adachi, P. J. C., Hodson, G., Willoughby, T., & Zanette, S. (2015). Brothers and Sisters in Arms: Intergroup Cooperation in a Violent Shooter Game Can Reduce Intergroup Bias. *Psychology of Violence*, *5*(4), 455–462. https://doi.org/10.1037/a0037407

Allport, G. W. (1954). *The nature of prejudice*. Reading, Mass.: Addison-Wesley.

Alves, H., Koch, A., & Unkelbach, C. (2018). A Cognitive-Ecological Explanation of Intergroup Biases. *Psychological Science*, *29*(7), 1126–1133. https://doi.org/10.1177/0956797618756862

Anderson, N. H. (1981). *Foundations of information integration theory*. New York: Academic Press.

Aron, A., Aron, E. N., & Smollan, D. (1992). Inclusion of Other in the Self Scale and the structure of interpersonal closeness. *Journal of Personality and Social Psychology*, *63*(4), 596–612. https://doi.org/10.1037/0022-3514.63.4.596

Aronson, E. (2002). *The social animal* (8. ed.). New York: Worth Publishers.

Aronson, E., & Cope, V. (1968). My enemy's enemy is my friend. *Journal of Personality and Social Psychology*, *8*(1, Pt.1), 8–12. https://doi.org/10.1037/h0021234

Asch, S. E. (1952). *Social Psychology*. New York: Prentice-Hall.

Bar-Tal, D., Kruglanski, A. W., & Klar, Y. (1989). Conflict termination: An epistemological analysis of international cases. *Political Psychology*, *10*(2), 233–255.

Batson, C. D., Pate, S., Lawless, H., Sparkman, P., Lambers, S., & Worman, B. (1979). Helping Under Conditions of Common Threat: Increased "We-Feeling" or Ensuring Reciprocity. *Social Psychology Quarterly*, *42*(4), 410–414. https://doi.org/10.2307/3033812

Bernstein, M. J., Young, S. G., & Hugenberg, K. (2007). The cross-category effect: Mere social categorization is sufficient to elicit an own-group bias in face recognition. *Psychological Science*, *18*(8), 706–712.

Bhatia, S. (2017). The semantic representation of prejudice and stereotypes. *Cognition*, *164*, 46–60. https://doi.org/10.1016/j.cognition.2017.03.016

Bhatti, A., & Kimmich, D. (Eds.) (2015). *Ähnlichkeit: Ein kulturtheoretisches Paradigma*. Konstanz: Konstanz University Press.

Biernat, M., & Manis, M. (1994). Shifting standards and stereotype-based judgments. *Journal of Personality and Social Psychology*, *66*(1), 5–20. https://doi.org/10.1037/0022-3514.66.1.5

Bigler, R. S., & Liben, L. S. (2006). A developmental intergroup theory of social stereotypes and prejudice. In R. V. Kail (Ed.), *Advances in Child Development and Behavior* (Vol. 34, pp. 39–89). San Diego: Academic Press [Imprint]; Elsevier Science & Technology Books. https://doi.org/10.1016/S0065-2407(06)80004-2

Bigler, R. S., & Liben, L. S. (2007). Developmental Intergroup Theory. *Current Directions in Psychological Science*, *16*(3), 162–166. https://doi.org/10.1111/j.1467-8721.2007.00496.x

Bodenhausen, G. V., & Peery, D. (2009). Social Categorization and Stereotyping In vivo: The VUCA Challenge. *Social and Personality Psychology Compass*, *3*(2), 133–151.

Boghardt, L. P. (2014). Qatar and ISIS Funding: The U.S. Approach. Retrieved from https://www.washingtoninstitute.org/policy-analysis/view/qatar-and-isis-funding-the-u.s.-approach

Boldry, J. G., Gaertner, L., & Quinn, J. (2007). Measuring the Measures: A Meta-Analytic Investigation of the Measures of Outgroup Homogeneity. *Group Processes & Intergroup Relations*, *10*(2), 157–178. https://doi.org/10.1177/1368430207075153

Bonefeld, M., & Dickhäuser, O. (2018). (Biased) Grading of Students' Performance: Students' Names, Performance Level, and Implicit Attitudes. *Frontiers in Psychology*, *9*, 1–13. https://doi.org/10.3389/fpsyg.2018.00481

Bosson, J., Johnson, A. B., Niederhoffer, K., & Swann, W. B. (2006). Interpersonal chemistry through negativity: Bonding by sharing negative attitudes about others. *Personal Relationships*, *13*(2), 135–150. https://doi.org/10.1111/j.1475-6811.2006.00109.x

Boudon, R., & Silverman, D. (1981). *The logic of social action: An introduction to sociological analysis*. London: Routledge & Kegan Paul.

Branscombe, N. R., Ellemers, N., Spears, R., Doosje, B., & others (1999). The context and content of social identity threat. In N. Ellemers, R. Spears, & B. Doosje (Ed.), *Social identity: Context, commitment, content* (pp. 35–58). Oxford, England: Blackwell Science.

Brewer, M. B. (1988). A dual process model of impression formation. In *Advances in social cognition, Vol. 1. A dual process model of impression formation* (pp. 1–36). Hillsdale, NJ, US: Lawrence Erlbaum Associates, Inc.

Brewer, M. B., & Miller, N. (1984). *Groups in contact: The psychology of desegregation*: Academic Press. Retrieved from https://www.amazon.de/Advances-Social-Cognition-Impression-Formation/dp/0898596734#reader_0898596734

Brubaker, R. (2007). *Ethnizität ohne Gruppen* (G. Gockel & S. Schuhmacher, Trans.). Hamburg: Hamburger Edition. Retrieved from http://deposit.d-nb.de/cgi-bin/dokserv?id=2951906&prov=M&dok_var=1&dok_ext=htm

Bruhn, J. (2009). The Concept of Social Cohesion. In J. Bruhn (Ed.), *The Group Effect* (pp. 31–48). Boston, MA: Springer US. https://doi.org/10.1007/978-1-4419-0364-8_2

Bruner, J. S. (1957). On perceptual readiness. *Psychological Review*, *64*(2), 123–152. https://doi.org/10.1037/h0043805

Bruner, J. S., Goodnow, J. J., & Austin, G. A. (1956). *A study of thinking*. New York: Wiley.

Bryson, J. B., & Franco, L. B. (1976). Experimental differentiation of meaning change and averaging explanations for context effects. *Memory & Cognition*, *4*(3), 337–344. https://doi.org/10.3758/BF03213186

Burnstein, E., & McRae, A. V. (1962). Some effects of shared threat and prejudice in racially mixed groups. *The Journal of Abnormal and Social Psychology*, *64*(4), 257–263. https://doi.org/10.1037/h0046022

Cartwright, D., & Harary, F. (1956). Structural balance: a generalization of Heider's theory. *Psychological Review*, *63*(5), 277–293. https://doi.org/10.1037/h0046049

Chance, J. E., & Goldstein, A. G. (1996). The other-race effect and eyewitness identification. In *Psychological issues in eyewitness identification* (pp. 153–176). Hillsdale, NJ, US: Lawrence Erlbaum Associates, Inc.

Cole, E. R. (2009). Intersectionality and research in psychology. *The American Psychologist*, *64*(3), 170–180. https://doi.org/10.1037/a0014564

Collisson, B., & Howell, J. L. (2014). The liking-similarity effect: Perceptions of similarity as a function of liking. *The Journal of Social Psychology*, *154*(5), 384–400.

Correll, J., Park, B., & Allegra Smith, J. (2008). Colorblind and Multicultural Prejudice Reduction Strategies in High-Conflict Situations. *Group Processes & Intergroup Relations*, *11*(4), 471–491. https://doi.org/10.1177/1368430208095401

Darley, J. M., & Gross, P. H. (1983). A hypothesis-confirming bias in labeling effects. *Journal of Personality and Social Psychology*, *44*(1), 20–33.

Degner, J., Imhoff, R., & Dunham, Y. (2020). *Assessing Categorical Person Construal in the Who-said-What paradigm: Can we separate Categorization from Perceptual Similarity?* Manuscript in preparation.

Deska, J. C. (2018). *They're all the same to me: Homogeneous groups are denied mind* (Doctoral dissertation). Miami University, Oxford, Ohio.

Deutsch, M. (1968). Field theory in social psychology. In E. Aronson & G. Lindzey (Eds.), *The handbook of social psychology* (2nd ed., Vol. 1, pp. 412–487). Menlo Park, CA: Addison-Wesley Publishing Company.

Dotsch, R., Wigboldus, D. H. J., & van Knippenberg, A. (2011). Biased allocation of faces to social categories. *Journal of Personality and Social Psychology*, *100*(6), 999–1014. https://doi.org/10.1037/a0023026

Dovidio, J. F., Gaertner, S. L., Hodson, G., & Houlette, M. A. (2004). Social inclusion and exclusion: Recategorization and the perception of intergroup boundaries. In D. Abrams, M. A. Hogg, & J. M. Marques (Eds.), *Social psychology of inclusion and exclusion* (pp. 263–282). Psychology Press.

Dovidio, J. F., & Morris, W. N. (1975). Effects of stress and commonality of fate on helping behavior. *Journal of Personality and Social Psychology*, *31*(1), 145–149.

Drake, N. (2018). They Saw Earth From Space. Here's How It Changed Them. Retrieved from https://www.nationalgeographic.com/magazine/2018/03/astronauts-space-earth-perspective/

Drury, J., Cocking, C., Reicher, S. [Steve], Burton, A., Schofield, D., Hardwick, A., . . . Langston, P. (2009). Cooperation versus competition in a mass emergency evacuation: A new laboratory simulation and a new theoretical model. *Behavior Research Methods*, *41*(3), 957–970. https://doi.org/10.3758/BRM.41.3.957

Duncan, B. L. (1976). Differential social perception and attribution of intergroup violence: Testing the lower limits of stereotyping of blacks. *Journal of Personality and Social Psychology*, *34*(4), 590–598.

Durante, F., & Fiske, S. T. (2017). How social-class stereotypes maintain inequality. *Current Opinion in Psychology*, *18*, 43–48. https://doi.org/10.1016/j.copsyc.2017.07.033

Feingold, G. A. (1914). Influence of environment on identification of persons and things. *J. Am. Inst. Crim. L. & Criminology*, *5*, 39–51.

Feshbach, S., & Singer, R. (1957). The effects of personal and shared threats upon social prejudice. *Journal of Abnormal Psychology*, *54*(3), 411–416.

Fiedler, K. (2017). What Constitutes Strong Psychological Science? The (Neglected) Role of Diagnosticity and A Priori Theorizing. *Perspectives on Psychological Science: A Journal of the Association for Psychological Science*, *12*(1), 46–61. https://doi.org/10.1177/1745691616654458

Fiedler, K. (2018). The Creative Cycle and the Growth of Psychological Science. *Perspectives on Psychological Science : A Journal of the Association for Psychological Science*, *13*(4), 433–438. https://doi.org/10.1177/1745691617745651

Fiske, S. T., Cuddy, A. J. C., Glick, P., & Xu, J. (2002). A model of (often mixed) stereotype content: Competence and warmth respectively follow from perceived status and competition. *Journal of Personality and Social Psychology*, *82*(6), 878–902. https://doi.org/10.1037//0022-3514.82.6.878

Fiske, S. T., Lin, M., & Neuberg, S. L. (2018). The continuum model: Ten years later. In S. T. Fiske (Ed.), *World Library of Psychologists. Social Cognition: Selected Works of Susan Fiske* (1st ed., pp. 231–254). Milton: Taylor and Francis.

Fiske, S. T., & Neuberg, S. L. (1990). A Continuum of Impression Formation, from Category-Based to Individuating Processes: Influences of Information and Motivation on Attention and Interpretation. In *Advances in Experimental Social Psychology* (Vol. 23, pp. 1–74). Elsevier. https://doi.org/10.1016/S0065-2601(08)60317-2

Flade, F., Klar, Y., & Imhoff, R. (2019). Unite against: A common threat invokes spontaneous decategorization between social categories. *Journal of Experimental Social Psychology*, *85*, 103890. https://doi.org/10.1016/j.jesp.2019.103890

Freeman, J. B., Stolier, R. M., Brooks, J. A., & Stillerman, B. A. (2018). The neural representational geometry of social perception. *Current Opinion in Psychology*, *24*, 83–91. https://doi.org/10.1016/j.copsyc.2018.10.003

Gaertner, S. L., Dovidio, J. F., Banker, B. S., Houlette, M. A., Johnson, K. M., & McGlynn, E. A. (2000). Reducing intergroup conflict: From superordinate goals to decategorization, recategorization, and mutual differentiation. *Group Dynamics: Theory, Research, and Practice*, *4*(1), 98–114. https://doi.org/10.1037//1089-2699.4.1.98

Gentner, D., & Markman, A. B. (1994). Structural alignment in comparison: No difference without similarity. *Psychological Science*, *5*(3), 152–158.

Goffman, E. (1959). *The presentation of self in everyday life*. *Anchor books*. New York, NY: Doubleday.

Goh, J. X., Hall, J. A., & Rosenthal, R. (2016). Mini Meta-Analysis of Your Own Studies: Some Arguments on Why and a Primer on How. *Social and Personality Psychology Compass*, *10*(10), 535–549. https://doi.org/10.1111/spc3.12267

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics* (Vol. 1): Wiley New York.

Greitemeyer, T., Traut-Mattausch, E., & Osswald, S. (2012). How to ameliorate negative effects of violent video games on cooperation: Play it cooperatively in a team. *Computers in Human Behavior*, *28*(4), 1465–1470. https://doi.org/10.1016/j.chb.2012.03.009

Grice, H. P. (1975). Logic and Conversation. *Syntax and Semantics*. (3), 41-58.

Groebner, V. (2003). Haben Hautfarben eine Geschichte? Personenbeschreibungen und ihre Kategorien zwischen dem 13. und dem 16. Jahrhundert. *Zeitschrift Für Historische Forschung*, *30*(1), 1–17. Retrieved from http://www.jstor.org/stable/43572114

Hahn, A., Banchefsky, S., Park, B., & Judd, C. M. (2015). Measuring intergroup ideologies: Positive and negative aspects of emphasizing versus looking beyond group differences. *Personality & Social Psychology Bulletin*, *41*(12), 1646–1664. https://doi.org/10.1177/0146167215607351

Hahn, U., Chater, N., & Richardson, L. B. (2003). Similarity as transformation. *Cognition*, *87*(1), 1–32. https://doi.org/10.1016/S0010-0277(02)00184-1

Haines, E. L., Deaux, K., & Lofaro, N. (2016). The Times They Are a-Changing … or Are They Not? A Comparison of Gender Stereotypes, 1983–2014. *Psychology of Women Quarterly*, *40*(3), 353–363. https://doi.org/10.1177/0361684316634081

Harari, H., & McDavid, J. W. (1973). Name stereotypes and teachers' expectations. *Journal of Educational Psychology*, *65*(2), 222–225. https://doi.org/10.1037/h0034978

Haslam, S. A., Turner, J. C., Oakes, P. J., McGarty, C., & Hayes, B. K. (1992). Context-dependent variation in social stereotyping 1: The effects of intergroup relations as mediated by social change and frame of reference. *European Journal of Social Psychology*, *22*(1), 3–20. https://doi.org/10.1002/ejsp.2420220104

Hayden, S. R., Jackson, T. T., & Guydish, J. (1984). Helping behavior of females: Effects of stress and commonality of fate. *The Journal of Psychology: Interdisciplinary and Applied*, *117*(2), 233–237. https://doi.org/10.1080/00223980.1984.9923683

Heck, D. W., Arnold, N. R., & Arnold, D. (2018). Treebugs: An R package for hierarchical multinomial-processing-tree modeling. *Behavior Research Methods*, *50*(1), 264–284. https://doi.org/10.3758/s13428-017-0869-7

Heider, F. (1946). Attitudes and cognitive organization. *The Journal of Psychology*, *21*(1), 107–112.

Heine, S. J., Proulx, T., & Vohs, K. D. (2006). The meaning maintenance model: On the coherence of social motivations. *Personality and Social Psychology Review : An Official Journal of the Society for Personality and Social Psychology, Inc*, *10*(2), 88–110. https://doi.org/10.1207/s15327957pspr1002_1

Henry, P. J., & Sears, D. O. (2002). The Symbolic Racism 2000 Scale. *Political Psychology*, *23*(2), 253–283.

Hewstone, M., & Brown, R. (1986). Contact is not enough: An intergroup perspective on the 'contact hypothesis.'. In *Social psychology and society. Contact and conflict in intergroup encounters* (pp. 1–44). Cambridge, MA, US: Basil Blackwell.

Hewstone, M., Rubin, M., & Willis, H. (2002). Intergroup bias. *Annual Review of Psychology*, *53*, 575–604. https://doi.org/10.1146/annurev.psych.53.100901.135109

Hirschauer, S. (2014). Un/doing Differences. Die Kontingenz sozialer Zugehörigkeiten / Un/doing Differences. The Contingency of Social Belonging. *Zeitschrift Für Soziologie*, *43*(3), 170–191.

Hirschberger, G., Ein-Dor, T., Leidner, B., & Saguy, T. (2016). How Is Existential Threat Related to Intergroup Conflict? Introducing the Multidimensional Existential Threat (MET) Model. *Frontiers in Psychology*, *7*, 1–18. https://doi.org/10.3389/fpsyg.2016.01877

Hong, Y.-y., Coleman, J., Chan, G., Wong, R. Y. M., Chiu, C.-y., Hansen, I. G., . . . Fu, H.-y. (2004). Predicting intergroup bias: The interactive effects of implicit theory and social identity. *Personality & Social Psychology Bulletin*, *30*(8), 1035–1047. https://doi.org/10.1177/0146167204264791

Hugenberg, K., & Bodenhausen, G. V. (2004). Ambiguity in Social Categorization: The Role of Prejudice and Facial Affect in Race Categorization. *Psychological Science*, *15*(5), 342–345. Retrieved from https://doi.org/10.1111/j.0956-7976.2004.00680.x

Hugenberg, K., Miller, J., & Claypool, H. M. (2007). Categorization and individuation in the cross-race recognition deficit: Toward a solution to an insidious problem. *Journal of Experimental Social Psychology*, *43*(2), 334–340. https://doi.org/10.1016/j.jesp.2006.02.010

Hugenberg, K., & Sacco, D. F. (2008). Social categorization and stereotyping: How social categorization biases person perception and face memory. *Social and Personality Psychology Compass*, *2*(2), 1052–1072.

Hugenberg, K., Young, S. G., Bernstein, M. J., & Sacco, D. F. (2010). The categorization-individuation model: An integrative account of the other-race recognition deficit. *Psychological Review*, *117*(4), 1168–1187. https://doi.org/10.1037/a0020463

Imhoff, R., & Koch, A. (2017). How Orthogonal Are the Big Two of Social Perception? On the Curvilinear Relation Between Agency and Communion. *Perspectives on Psychological Science : A Journal of the Association for Psychological Science*, *12*(1), 122–137. https://doi.org/10.1177/1745691616657334

Imhoff, R., Koch, A., & Flade, F. (2018). (Pre)occupations: A data-driven model of jobs and its consequences for categorization and evaluation. *Journal of Experimental Social Psychology*, *77*, 76–88. https://doi.org/10.1016/j.jesp.2018.04.001

Imhoff, R., Woelki, J., Hanke, S., & Dotsch, R. (2013). Warmth and competence in your face! Visual encoding of stereotype content. *Frontiers in Psychology*, *4*, 386. https://doi.org/10.3389/fpsyg.2013.00386

Ito, T. A., & Urland, G. R. (2003). Race and gender on the brain: Electrocortical measures of attention to the race and gender of multiply categorizable individuals. *Journal of Personality and Social Psychology*, *85*(4), 616–626. https://doi.org/10.1037/0022-3514.85.4.616

Job, V., Dweck, C. S., & Walton, G. M. (2010). Ego depletion--is it all in your head? Implicit theories about willpower affect self-regulation. *Psychological Science*, *21*(11), 1686–1693. https://doi.org/10.1177/0956797610384745

Johnson, K. L., Lick, D. J., & Carpinella, C. M. (2015). Emergent Research in Social Vision: An Integrated Approach to the Determinants and Consequences of Social Categorization. *Social and Personality Psychology Compass*, *9*(1), 15–30. https://doi.org/10.1111/spc3.12147

Kawakami, K., Amodio, D. M., & Hugenberg, K. (2017). Intergroup perception and cognition: An integrative framework for understanding the causes and consequences of

social categorization. In *Advances in experimental social psychology* (Vol. 55, pp. 1–80). Elsevier.

Kearns, E. M., Betus, A. E., & Lemieux, A. F. (2019). Why Do Some Terrorist Attacks Receive More Media Attention Than Others? *Justice Quarterly*, *36*(6), 985–1022. https://doi.org/10.1080/07418825.2018.1524507

Kelman, H. C. (1999). The Interdependence of Israeli and Palestinian National Identities: The Role of the Other in Existential Conflicts. *Journal of Social Issues*, *55*(3), 581–600. https://doi.org/10.1111/0022-4537.00134

Klapper, A., Dotsch, R., van Rooij, I., & Wigboldus, D. H. J. (2016). Do we spontaneously form stable trustworthiness impressions from facial appearance? *Journal of Personality and Social Psychology*, *111*(5), 655–664. https://doi.org/10.1037/pspa0000062

Klapper, A., Dotsch, R., van Rooij, I., & Wigboldus, D. H. J. (2017). Four meanings of "categorization": A conceptual analysis of research on person perception. *Social and Personality Psychology Compass*, *11*(8), 1–16. https://doi.org/10.1111/spc3.12336

Klauer, K. C. (2010). Hierarchical Multinomial Processing Tree Models: A Latent-Trait Approach. *Psychometrika*, *75*(1), 70–98. https://doi.org/10.1007/s11336-009-9141-0

Klauer, K. C., & Ehrenberg, K. (2005). Social categorization and fit detection under cognitive load: Efficient or effortful? *European Journal of Social Psychology*, *35*(4), 493–516. https://doi.org/10.1002/ejsp.266

Klauer, K. C., Hölzenbein, F., Calanchini, J., & Sherman, J. W. (2014). How malleable is categorization by race? Evidence for competitive category use in social categorization. *Journal of Personality and Social Psychology*, *107*(1), 21–40. https://doi.org/10.1037/a0036609

Klauer, K. C., & Wegener, I. (1998). Unraveling social categorization in the "who said what?" paradigm. *Journal of Personality and Social Psychology*, *75*(5), 1155–1178.

Koch, A., Imhoff, R., Dotsch, R., Unkelbach, C., & Alves, H. (2016). The ABC of stereotypes about groups: Agency/socioeconomic success, conservative-progressive beliefs, and communion. *Journal of Personality and Social Psychology*, *110*(5), 675–709. https://doi.org/10.1037/pspa0000046

Kunda, Z., & Sherman-Williams, B. (1993). Stereotypes and the Construal of Individuating Information. *Personality & Social Psychology Bulletin*, *19*(1), 90–99. https://doi.org/10.1177/0146167293191010

Kunda, Z., & Thagard, P. (1996). Forming impressions from stereotypes, traits, and behaviors: A parallel-constraint-satisfaction theory. *Psychological Review*, *103*(2), 284–308. https://doi.org/10.1037/0033-295X.103.2.284

Kurzban, R., Tooby, J., & Cosmides, L. (2001). Can race be erased? Coalitional computation and social categorization. *Proceedings of the National Academy of Sciences*, *98*(26), 15387–15392.

Laqueur, T. (2003). *Making sex: Body and gender from the Greeks to Freud* (10. print). Cambridge, Mass.: Harvard University Press.

Leeuw, J. R. de, Andrews, J. K., Livingston, K. R., & Chin, B. M. (2016). The Effects of Categorization on Perceptual Judgment are Robust across Different Assessment Tasks. *Collabra*, *2*(1), 1–9. https://doi.org/10.1525/collabra.32

SoSci Survey [Computer software] (2014). Retrieved from https://www.soscisurvey.de

Leonardelli, G. J., & Toh, S. M. (2015). Social Categorization in Intergroup Contexts: Three Kinds of Self-Categorization. *Social and Personality Psychology Compass*, *9*(2), 69–87. https://doi.org/10.1111/spc3.12150

Levin, D. T. (1996). Classifying faces by race: The structure of face categories. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*(6), 1364–1382. https://doi.org/10.1037/0278-7393.22.6.1364

Lewin, K. (1943). Defining the 'field at a given time.'. *Psychological Review*, *50*(3), 292–310. https://doi.org/10.1037/h0062738

Liberman, Z., Woodward, A. L., & Kinzler, K. D. (2017). The Origins of Social Categorization. *Trends in Cognitive Sciences*, *21*(7), 556–568. https://doi.org/10.1016/j.tics.2017.04.004

Ma, D. S., Correll, J., & Wittenbrink, B. (2015). The Chicago face database: A free stimulus set of faces and norming data. *Behavior Research Methods*, *47*(4), 1122–1135. https://doi.org/10.3758/s13428-014-0532-5

Maass, A. (1999). Linguistic intergroup bias: Stereotype perpetuation through language. *Advances in Experimental Social Psychology*, *31*, 79–122.

MacLin, O. H., & Malpass, R. S. (2001). Racial categorization of faces: The ambiguous race face effect. *Psychology, Public Policy, and Law*, *7*(1), 98–118. https://doi.org/10.1037/1076-8971.7.1.98

Macrae, C. N., Stangor, C., & Milne, A. B. (1994). Activating Social Stereotypes: A Functional Analysis. *Journal of Experimental Social Psychology*, *30*(4), 370–389. https://doi.org/10.1006/jesp.1994.1018

Malpass, R. S., & Kravitz, J. (1969). Recognition for faces of own and other race. *Journal of Personality and Social Psychology*, *13*(4), 330–334. https://doi.org/10.1037/h0028434

Malt, B. C., Sloman, S. A., & Gennari, S. P. (2003). Universality and language specificity in object naming. *Journal of Memory and Language*, *49*(1), 20–42. https://doi.org/10.1016/S0749-596X(03)00021-4

Markant, J., & Scott, L. S. (2017). Attention and Perceptual Learning Interact in the Development of the Other-Race Effect. *Current Directions in Psychological Science*, *27*(3), 163–169. https://doi.org/10.1177/0963721418769884

Markovsky, B. (2018). *Theory Construction & Analysis: Ten Lectures.* EASP Summer School, Zurich.

McGarty, C. (1999). *Categorization in Social Psychology*. London: SAGE Publications. Retrieved from http://site.ebrary.com/lib/alltitles/docDetail.action?docID=10567045

Meenes, M. (1943). A Comparison of Racial Stereotypes of 1935 and 1942. *The Journal of Social Psychology*, *17*(2), 327–336. https://doi.org/10.1080/00224545.1943.9712287

Meissner, C. A., & Brigham, J. C. (2001). Thirty years of investigating the own-race bias in memory for faces: A meta-analytic review. *Psychology, Public Policy, and Law*, *7*(1), 3–35. https://doi.org/10.1037//1076-8971.7.1.3

Miller, C. T. (1988). Categorization and the physical attractiveness stereotype. *Social Cognition*, *6*(3), 231–251.

Minear, M., & Park, D. C. (2004). A lifespan database of adult facial stimuli. *Behavior Research Methods, Instruments, & Computers: A Journal of the Psychonomic Society, Inc*, *36*(4), 630–633. https://doi.org/10.3758/bf03206543

Molenberghs, P., & Morrison, S. (2014). The role of the medial prefrontal cortex in social categorization. *Social Cognitive and Affective Neuroscience*, *9*(3), 292–296. https://doi.org/10.1093/scan/nss135

Mussweiler, T. (2003). 'Everything is relative': Comparison processes in social judgment The 2002 Jaspars Lecture. *European Journal of Social Psychology*, *33*(6), 719–733. https://doi.org/10.1002/ejsp.169

Natu, V., Raboy, D., & O'Toole, A. J. (2011). Neural correlates of own- and other-race face perception: Spatial and temporal response differences. *NeuroImage*, *54*(3), 2547–2555. https://doi.org/10.1016/j.neuroimage.2010.10.006

North, M. S., & Fiske, S. T. (2012). An inconvenienced youth? Ageism and its potential intergenerational roots. *Psychological Bulletin*, *138*(5), 982–997. https://doi.org/10.1037/a0027843

Oakes, P. J. (1987). The salience of social categories. In J. C. Turner, M. A. Hogg, P. J. Oakes, S. D. Reicher, & M. S. Wetherell (Eds.), *Rediscovering the social group: A self-categorization theory* (pp. 117–141). Basil Blackwell.

Oakes, P. J. (2008). The Root of All Evil in Intergroup Relations? Unearthing the Categorization Process. In R. Brown & S. L. Gaertner (Eds.), *Blackwell handbook of social psychology. Blackwell handbook of social psychology: Intergroup processes* (pp. 3–21). Malden, MA: Blackwell. https://doi.org/10.1002/9780470693421.ch1

Oakes, P. J., Haslam, S. A., & Turner, J. C. (1994). *Stereotyping and social reality*: Blackwell Publishing.

Oeberst, A., & Imhoff, R. (2020). *Towards parsimony in bias research. Integrating disparate research lines in a common framework of belief-consistent information processing.* Manuscript in preparation.

Ohmann, K., & Burgmer, P. (2016). Nothing compares to me: How narcissism shapes comparative thinking. *Personality and Individual Differences*, *98*, 162–170. https://doi.org/10.1016/j.paid.2016.03.069

Otten, S. (2016). The Minimal Group Paradigm and its maximal impact in research on social categorization. *Current Opinion in Psychology*, *11*, 85–89. https://doi.org/10.1016/j.copsyc.2016.06.010

Park, B., & Judd, C. M. (2005). Rethinking the link between categorization and prejudice within the social cognition perspective. *Personality and Social Psychology Review*, *9*(2), 108–130. https://doi.org/10.1207/s15327957pspr0902_2

Park, B., & Rothbart, M. (1982). Perception of out-group homogeneity and levels of social categorization: Memory for the subordinate attributes of in-group and out-group members.

Journal of Personality and Social Psychology, *42*(6), 1051–1068. https://doi.org/10.1037/0022-3514.42.6.1051

Pepitone, A., & Kleiner, R. (1957). The effects of threat and frustration on group cohesiveness. *The Journal of Abnormal and Social Psychology*, *54*(2), 192–199. https://doi.org/10.1037/h0049040

Pettigrew, T. F. (1998). Intergroup contact theory. *Annual Review of Psychology*, *49*, 65–85. https://doi.org/10.1146/annurev.psych.49.1.65

Pexman, P. M., & Olineck, K. M. (2002). Understanding irony: How do stereotypes cue speaker intent? *Journal of Language and Social Psychology*, *21*(3), 245–274.

Pietraszewski, D., & Schwartz, A. (2014). Evidence that accent is a dimension of social categorization, not a byproduct of perceptual salience, familiarity, or ease-of-processing. *Evolution and Human Behavior*, *35*(1), 43–50.

Rakić, T., Steffens, M. C., & Mummendey, A. (2011). Blinded by the accent! The minor role of looks in ethnic categorization. *Journal of Personality and Social Psychology*, *100*(1), 16–29. https://doi.org/10.1037/a0021522

Rees, H. R., Ma, D. S., & Sherman, J. W. (2020). Examining the Relationships Among Categorization, Stereotype Activation, and Stereotype Application. *Personality & Social Psychology Bulletin*, *46*(4), 499–513. https://doi.org/10.1177/0146167219861431

Riek, B. M., Mania, E. W., & Gaertner, S. L. (2006). Intergroup threat and outgroup attitudes: A meta-analytic review. *Personality and Social Psychology Review : An Official Journal of the Society for Personality and Social Psychology, Inc*, *10*(4), 336–353.

Rosch, E. (1978). Principles of Categorization. In E. Rosch & B. Lloyd (Eds.), *Cognition and categorization* (pp. 27–48). Hillsdale, NJ: Lawrence Erlbaum.

Rosenthal, R., & Jacobson, L. (1968). Pygmalion in the classroom. *The Urban Review*, *3*(1), 16–20. https://doi.org/10.1007/BF02322211

Rouhana, N. N., & Bar-Tal, D. (1998). Psychological dynamics of intractable ethnonational conflicts: The Israeli–Palestinian case. *American Psychologist*, *53*(7), 761–770. https://doi.org/10.1037/0003-066X.53.7.761

Sagar, H. A., & Schofield, J. W. (1980). Racial and behavioral cues in Black and White children's perceptions of ambiguously aggressive acts. *Journal of Personality and Social Psychology*, *39*(4), 590–598. https://doi.org/10.1037/0022-3514.39.4.590

Sasaki, S. J., & Vorauer, J. D. (2013). Ignoring Versus Exploring Differences Between Groups: Effects of Salient Color-Blindness and Multiculturalism on Intergroup Attitudes and Behavior. *Social and Personality Psychology Compass*, *7*(4), 246–259. https://doi.org/10.1111/spc3.12021

Scherr, A. (2020). Soziale Distanz und Diskriminierung. In A. Röder & D. Zifonun (Eds.), *Handbuch Migrationssoziologie* (pp. 1–32). Wiesbaden: Springer Fachmedien Wiesbaden. https://doi.org/10.1007/978-3-658-20773-1_27-1

Schubert, T. W., & Otten, S. (2002). Overlap of Self, Ingroup, and Outgroup: Pictorial Measures of Self-Categorization. *Self and Identity*, *1*(4), 353–376. https://doi.org/10.1080/152988602760328012

Seago, D. W. (1947). Stereotypes: Before Pearl Harbor and after. *The Journal of Psychology*, *23*(1), 55–63. https://doi.org/10.1080/00223980.1947.9917320

Semin, G. R., & Fiedler, K. (1991). The linguistic category model, its bases, applications and range. *European Review of Social Psychology*, *2*(1), 1–30.

Sharpe, D., & Whelton, W. J. (2016). Frightened by an old scarecrow: The remarkable resilience of demand characteristics. *Review of General Psychology*, *20*(4), 349–368. https://doi.org/10.1037/gpr0000087

Sherif, M. (1958). Superordinate Goals in the Reduction of Intergroup Conflict. *American Journal of Sociology*, *63*(4), 349–356. Retrieved from www.jstor.org/stable/2774135

Sherif, M., & Hovland, C. I. (1961). *Social judgment: Assimilation and contrast effects in communication and attitude change*. *Yale Studies in attitude and communication: Vol. 4*. Westport, Conn.: Greenwood Press.

Sherif, M., & Sherif, C. W. (1953). *Groups in harmony and tension; an integration of studies of intergroup relations*. Oxford, England: Harper & Brothers.

Sherman, J. W., Klein, S. B., Laskey, A., & Wyer, N. A. (1998). Intergroup Bias in Group Judgment Processes: The Role of Behavioral Memories. *Journal of Experimental Social Psychology*, *34*(1), 51–65. https://doi.org/10.1006/jesp.1997.1342

Sherman, J. W., Macrae, C. N., & Bodenhausen, G. V. (2011). Attention and Stereotyping: Cognitive Constraints on the Construction of Meaningful Social Impressions. *European Review of Social Psychology*, *11*(1), 145–175. https://doi.org/10.1080/14792772043000022

Simon, B., Hastedt, C., & Aufderheide, B. (1997). When self-categorization makes sense: The role of meaningful social categorization in minority and majority members' self-perception. *Journal of Personality and Social Psychology*, *73*(2), 310–320. https://doi.org/10.1037/0022-3514.73.2.310

Skowronski, J. J., & Carlston, D. E. (1989). Negativity and extremity biases in impression formation: A review of explanations. *Psychological Bulletin*, *105*(1), 131–142.

Slusher, M. P., & Anderson, C. A. (1987). When reality monitoring fails: The role of imagination in stereotype maintenance. *Journal of Personality and Social Psychology*, *52*(4), 653–662.

Stouffer, S. A. (1949). *The American soldier: 2. Combat and its aftermath*. *Studies in social psychology in World War II*. Princeton, NJ.

Sui, M., Dunaway, J., Sobek, D., Abad, A., Goodman, L., & Saha, P. (2017). U.S. News Coverage of Global Terrorist Incidents. *Mass Communication and Society*, *20*(6), 895–908. https://doi.org/10.1080/15205436.2017.1350716

Tajfel, H. (1970). Experiments in Intergroup Discrimination. *Scientific American*, *223*(5), 96–103.

Tajfel, H. (1972). La catégorisation sociale. In S. Moscovici (Ed.), *Sciences humaines et sociales. Introduction à la psychologie sociale* (Vol. 1, pp. 272–302). Paris: Larousse.

Tajfel, H., Billig, M. G., Bundy, R. P., & Flament, C. (1971). Social categorization and intergroup behaviour. *European Journal of Social Psychology*, *1*(2), 149–178.

Tajfel, H., & Wilkes, A. L. (1963). Classification and quantitative judgement. *British Journal of Psychology*, *54*(2), 101–114.

Taylor, J. B., Zurcher, L. A., & Key, W. H. (1970). *Tornado: A Community Responds to Disaster*. Seattle: University of Washington Press.

Taylor, S. E., Fiske, S. T., Etcoff, N. L., & Ruderman, A. J. (1978). Categorical and contextual bases of person memory and stereotyping. *Journal of Personality and Social Psychology*, *36*(7), 778–793. https://doi.org/10.1037/0022-3514.36.7.778

Tedeschi, J. T. (1981). *Impression Management Theory and Social Psychological Research*. Saint Louis: Elsevier Science.

Thibaut, J.-P., Dupont, M., & Anselme, P. (2002). Dissociations between categorization and similarity judgments as a result of learning feature distributions. *Memory & Cognition*, *30*(4), 647–656. https://doi.org/10.3758/bf03194966

Turner, J. C., Hogg, M. A., Oakes, P. J., Reicher, S. D., & Wetherell, M. S. (Eds.) (1987). *Rediscovering the social group: A self-categorization theory*: Basil Blackwell.

Turner, J. C., Oakes, P. J., Haslam, S. A., & McGarty, C. (1994). Self and collective: Cognition and social context. *Personality and Social Psychology Bulletin*, *20*(5), 454–463.

Uleman, J. S., Adil Saribay, S., & Gonzalez, C. M. (2008). Spontaneous inferences, implicit impressions, and implicit theories. *Annual Review of Psychology*, *59*, 329–360. https://doi.org/10.1146/annurev.psych.59.103006.093707

Unkelbach, C., Fiedler, K., Bayer, M., Stegmüller, M., & Danner, D. (2008). Why positive information is processed faster: The density hypothesis. *Journal of Personality and Social Psychology*, *95*(1), 36–49. https://doi.org/10.1037/0022-3514.95.1.36

Van Leeuwen, E., & Zagefka, H. (2017). *Intergroup Helping*. Cham: Springer International Publishing. Retrieved from http://search.ebscohost.com/login.aspx?direct=true&scope=site&db=nlebk&AN=1520215

Verhaeghen, P., Aikman, S. N., & van Gulick, A. E. (2011). Prime and prejudice: Co-occurrence in the culture as a source of automatic stereotype priming. *The British Journal of Social Psychology*, *50*(3), 501–518. https://doi.org/10.1348/014466610X524254

Vezzali, L., Drury, J., Versari, A., & Cadamuro, A. (2016). Sharing distress increases helping and contact intentions via social identification and inclusion of the other in the self: Children's prosocial behaviour after an earthquake. *Group Processes & Intergroup Relations*, *19*(3), 314–327. https://doi.org/10.1177/1368430215590492

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*, *36*(3), 1–48. Retrieved from http://www.jstatsoft.org/v36/i03/

Vosgerau, J., Simonsohn, U., Nelson, L. D., & Simmons, J. P. (2018, January 31). Internal Meta-Analysis Makes False-Positives Easier To Produce and Harder To Correct. Retrieved from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3271372

Wegener, I., & Klauer, K. C. (2005). Social Categorization Without Fit. *Zeitschrift Für Sozialpsychologie*, *36*(2), 91–101. https://doi.org/10.1024/0044-3514.36.2.91

Weisman, K., Johnson, M. V., & Shutts, K. (2015). Young children's automatic encoding of social categories. *Developmental Science*, *18*(6), 1036–1043. https://doi.org/10.1111/desc.12269

Werner, P. D., & LaRussa, G. W. (1985). Persistence and change in sex-role stereotypes. *Sex Roles*, *12*(9-10), 1089–1100. https://doi.org/10.1007/BF00288107

Wilder, D. A. (1990). Some determinants of the persuasive power of in-groups and out-groups: Organization of information and attribution of independence. *Journal of Personality and Social Psychology*, *59*(6), 1202–1213.

Wright, M. E. (1943). The influence of frustration upon the social relations of young children. *Journal of Personality*, *12*(2), 111–122. https://doi.org/10.1111/j.1467-6494.1943.tb01951.x

Yang, X., & Dunham, Y. (2019). Hard to disrupt: Categorization and enumeration by gender and race from mixed displays. *Journal of Experimental Social Psychology*, *85*, 103893. https://doi.org/10.1016/j.jesp.2019.103893

Young, S. G., Hugenberg, K., Bernstein, M. J., & Sacco, D. F. (2012). Perception and motivation in face recognition: A critical review of theories of the Cross-Race Effect. *Personality and Social Psychology Review : An Official Journal of the Society for Personality and Social Psychology, Inc*, *16*(2), 116–142. https://doi.org/10.1177/1088868311418987

Zajonc, R. B. (1968). Attitudinal effects of mere exposure. *Journal of Personality and Social Psychology*, *9*(2, Pt.2), 1–27.

Zerubavel, E. (1996). Lumping and splitting: Notes on social classification. In *Sociological Forum.* Symposium conducted at the meeting of Springer.

ERKLÄRUNG

gemäß § 6 Absatz 2 g) und gemäß § 6 Absatz 2 h) der Promotionsordnung der Fachbereiche 02, 05, 06, 07, 09 und 10 vom 04. April 2016

Name (ggf. Geburtsname): Flade

Vorname: Felicitas

Hiermit erkläre ich, dass ich die eingereichte Dissertation selbständig, ohne fremde Hilfe verfasst und mit keinen anderen als den darin angegebenen Hilfsmitteln angefertigt habe, dass die wörtlichen oder dem Inhalt nach aus fremden Arbeiten entnommenen Stellen, Zeichnungen, Skizzen, bildlichen Darstellungen und dergleichen als solche genau kenntlich gemacht sind.

Von der Ordnung zur Sicherung guter wissenschaftlicher Praxis in Forschung und Lehre und zum Verfahren zum Umgang mit wissenschaftlichem Fehlverhalten habe ich Kenntnis genommen.

Bei einer publikationsbasierten Promotion:
Meine Erklärung bezieht sich auf Schriften, die ich als alleiniger Autor bzw. Autorin eingereicht habe oder bei Ko-Autorenschaft auf jene Teile, für die ich mich verantwortlich zeichne.

Ich habe keine Hilfe von kommerziellen Promotionsberatern in Anspruch genommen.

14.04.2020
Datum

Unterschrift