Available online at www.sciencedirect.com

SCIENCE DIRECT®

ACADEMIC PRESS

Journal of Experimental Social Psychology 39 (2003) 332–354

Journal of Experimental Social Psychology

www.elsevier.com/locate/jesp

Crossed categorization and stereotyping: Structural analyses, effect patterns, and dissociative effects of context relevance☆

Karl Christoph Klauer,[*,1] Katja Ehrenberg,[1,2] and Ingo Wegener[3]

Psychologisches Institut, Rheinische Friedrich-Wilhelms-Universität Bonn, Römerstr. 164, 53117 Bonn, Germany

Received 10 May 2002; revised 10 September 2002

Abstract

Based on the "Who said what?" paradigm, a new method is proposed for investigating social categorization and non-evaluative stereotype application in crossed categorization. The method is applied in 5 studies that manipulated relative context relevance of crossed age and gender categories. Social categorization is characterized by 2 indices: *Relative subgroup memory* assesses the amount of subgroup formation represented in memory, and *relative category dominance* the relative weight of each dimension of categorization. Both indices were affected by context relevance. There was strong evidence for social categorization at the subgroup level, whereas stereotype application followed a simple pattern of category dominance, in which only the context-relevant dimension exerted an effect. The results bear on current models of category-based impression formation and intergroup perception, and on category-activation cum stereotype-inhibition models.

Every individual is a member of several social categories, which are associated with different stereotypical beliefs. Acknowledging this simple fact, a large amount of social psychological research has recently focused on questions of multiple category membership, exploring how multiple category memberships are represented in perceivers' minds, and how they relate to phenomena of stereotyping and prejudice. Most of the work addressing these questions has used so-called crossed categories in which two binary category distinctions are crossed orthogonally.

In their seminal work on crossed categorization, Deschamps (1977), Deschamps and Doise (1978), Doise (1978) argued that categorization leads to accentuation of differences between categories and exaggeration of similarities within categories (Tajfel, 1978). When two category dimensions cross-cut each other orthogonally, striving for accentuation on one dimension counteracts accentuation on the other. For example, in the presence of old and young men and women, accentuation along the age dimension requires assimilating men and women within each age group, whereas accentuation along the gender dimension requires assimilating old and young people within each gender group. In line with this category differentiation model, the loss in clarity of the category boundaries was found not only to reduce categorization of abstract stimuli (Deschamps, 1977; but see Crisp & Hewstone, 1999), but also to weaken the ascription of stereotypic traits as well as intergroup discrimination (Deschamps & Doise, 1978; Study 1 and 2, respectively).

Effect patterns in intergroup bias

A considerable amount of research has investigated intergroup bias when categories are crossed. From the

☆ The present work was supported by Grant No. Kl 614/7-2 from the Deutsche Forschungsgemeinschaft to the first author. We wish to thank Stefanie Birrenbach and Gabriela Munoz for collecting the data of Exps. 2 and 4, respectively, and Peter Freytag for his assistance in acquiring participants for Exp.1. Jan Mierke, Thorsten Meiser, and Arndt Bröder provided helpful feedback on a first draft of this manuscript.

* Corresponding author. Fax: +49-228-34229.

E-mail addresses: christoph.klauer@uni-bonn.de (K.C. Klauer), katja.ehrenberg@web.de (K. Ehrenberg).

1 These authors equally contributed to this paper.

2 Also corresponding author.

3 Present address: Department of Psychosomatic Medicine and Psychotherapy, University of Bonn, Germany.

0022-1031/03/$ - see front matter © 2003 Elsevier Science (USA). All rights reserved.
doi:10.1016/S0022-1031(03)00017-9

perceivers' perspective, target persons can be ingroup or outgroup members on each dimension of categorization considered separately, and the four subgroups defined by the category conjunctions have come to be denoted accordingly as double ingroup, partial ingroups, and double outgroup. One of the first systematic and widely cited classifications of possible effect patterns was proposed by Brewer, Ho, Lee, and Miller (1987) and further elaborated by Hewstone, Islam, and Judd (1993). One possibility, termed category dominance, is that one of the dimensions of categorization will dominate and the subordinate distinction will be ignored. Another possibility, termed additivity, is that there are additive and independent effects of both dimensions of classification. Category conjunction models predict that both dimensions of categorization interact so that a specific combination such as the double ingroup is evaluated differently from the other combinations which are treated more equally. Finally, hierarchical ordering also describes an interaction between the two dimensions of categorization in which the effects of one category distinction depend on categorization on the other dimension. This model predicts strong and pervasive ingroup–outgroup differentiation on the first, superordinate dimension. Ingroup–outgroup differentiation on the second category dimension is greater for target persons that are classified as ingroup members on the first category dimension than for targets that are members of the outgroup category on that dimension. Reviews of empirical studies have reported evidence for each of these patterns and most support for the additivity pattern (Crisp & Hewstone, 2000; Hewstone et al., 1993; Migdal, Hewstone, & Mullen, 1998; Urban & Miller, 1998). The reviews also suggest that the current body of evidence to some extent favors cognitive explanations of these patterns in terms of social categorization processes over motivational accounts as were derived from Social Identity Theory.

Nevertheless, comparatively little work has focused on the cognitive bases of dealing with a given person's multiple category memberships (Crisp, Hewstone, & Rubin, 2001; Macrae, Bodenhausen, & Milne, 1995) and on how these evaluative and behavioral effect patterns find their equivalents—and possibly their foundation—in patterns of social categorization (Stangor & Lange, 1993). Categorization has several effects on social perception (Zárate & Smith, 1990). There are evaluative and even discriminatory consequences as demonstrated in paradigms with arbitrary or minimal group distinctions (e.g., Brewer, 1979; Tajfel, Flament, Billig, & Bundy, 1971). Categorization also makes accessible stereotypes that are often seen as cognitive schemas and that serve a variety of functions over and above evaluative differentiation (e.g., Lippmann, 1922; Stangor & Lange, 1993) such as organizing traits and other characteristics that are linked to the category in long-term memory, directing attention to relevant features and providing expectancies on the stereotyped persons' traits and behaviors. The purpose of the present paper is to propose a new method that allows one to formulate and test analogues of the above models of possible effect patterns for a relatively direct and unobtrusive measure of social categorization and an unobtrusive measure of stereotype application. The method is applied in five studies that crossed gender categories (male versus female) with age categories (young versus old) while context relevance of these categories was manipulated between studies. The new method extends a multinomial model (Klauer & Wegener, 1998) for Taylor, Fiske, Etcoff, and Ruderman's (1978) "Who said what?" paradigm to the case of crossed categories.

The "Who said what?" paradigm

To measure processes of social categorization, the "Who said what?" paradigm is frequently used (van Knippenberg & Dijksterhuis, 2000), and the present method builds on that paradigm. Perceivers observe a discussion between members of two or more categories. In a recognition test, they are again shown the discussion statements and are then asked to assign each statement to its speaker. Participants' assignment errors can be classified into two kinds. Within-category errors occur if a statement is wrongly assigned to a person who is a member of the same category as the original speaker. Between-categories errors arise if a statement is assigned to a person from another category. The difference of the frequencies of the two kinds of error is often regarded as a measure of category salience in perceiving and interpreting the group discussion. Compared to other measures of group perception such as sorting tasks (e.g., Brewer & Lui, 1989) or ratings of typicality or similarity (e.g., Park, Wolsko, & Judd, 2001), it has the advantage of being relatively unobtrusive: participants never realize that social categorization is at issue and they are not asked to judge perceived group members, the group as a whole or relationships between group members. This renders a contamination by demand effects, active inferences, intrusions of different kinds of knowledge, and so forth unlikely. For example, in a recent comparison of different measures of subtyping, Park et al. (2001) found that a measure of memory confusions based on the "Who said what?" task predicted subtyping and pointed out that the confusion measure is relatively non-invasive in terms of preserving participants' spontaneous views of the group. The technique also allows one to use socially richer stimuli (e.g., group discussions) than is typically possible in the category verification task (Zárate & Smith, 1990) that has also been successfully used to assess social categorization in situations with multiple category memberships. Finally, as elaborated below, it provides a broader array of dependent measures than can be obtained through the use of analyses of clustering in recall (e.g., Hewstone, Macrae, Griffiths, Milne, & Brown, 1994).

A number of different processes contribute to participants' assignments of statements to speakers in the "Who said what?" paradigm, and Klauer and Wegener (1998) proposed to modify the paradigm slightly to disentangle their contributions. In the extended version, new statements or distracters are also presented in the assignment test, and participants first decide whether a given statement was presented as part of the discussion or whether it is new. In the former case they are then asked to assign the statement to its speaker as usual. This modification provides a richer data base, allowing one to disentangle the relative contributions of the different involved processes by means of a multinomial model of source discrimination (Batchelder & Riefer, 1999). In such a model, each latent cognitive process is assessed by a parameter reflecting the probability of the corresponding process to occur. The processes underlying the assignment behavior and accounted for in the multinomial model proposed by Klauer and Wegener can be described as follows:

Item discrimination reflects memory for the statements themselves, that is, the ability to identify an item as old or new. Identifying old statements correctly as old reflects recognition processes represented by the model parameter D. Distracters can be detected as new by means of so-called auto-noetic processes (Strack & Bless, 1994), which can be exemplified by inferences such as "If this statement had been presented, I would certainly remember it." The likelihood of distracter detection is captured by the model parameter D_N. This class of processes is affected, for example, by the similarity between old and new items (Klauer & Wegener, 1998, Exp. 2) or by cognitive load during encoding (Klauer & Wegener, 1998, Exp. 6).

Guessing of item status takes place if item memory fails such that participants cannot detect whether a given statement is old or new. The model parameter b reflects the probability to guess that an item is old rather than new, and is affected, for example, by the proportion of old items relative to distracters (Klauer & Wegener, 1998, Exp. 1).

Person memory refers to the ability to remember the speaker of a statement that is correctly recognized as old. This process is based on memory for the speaker's individuating features, and the corresponding model parameter c reflects the amount of individuation. Person memory is affected, for example, by the perceived similarity between discussants within each category (Klauer & Wegener, 1998, Exp. 5).

Person guessing. If the correct speaker cannot be remembered, there is still the possibility to choose him or her by chance if the correct category was remembered or guessed (see below). This option is accounted for by a fixed probability $1/n$, with n being the number of members within the respective category.

Category memory reflects the process of correctly remembering the speaker's social category membership, given that the statement is correctly recognized as old, but person memory fails. This process mirrors the strength of social categorization, and its contribution is assessed by the model parameter d. The amount of category memory is sensitive to, for example, a category's chronic or situational accessibility (Klauer & Wegener, 1998, Exp. 3).

Reconstructive category guessing refers to a possible bias to choose a speaker from a particular social category rather than another, if there is neither memory for the speaker nor for his or her category affiliation. It is captured by the model parameter a. Systematic tendencies to prefer one category over another under uncertainty can arise, for example, from stereotypic beliefs (Klauer, Wegener, & Ehrenberg, 2002) associating the content of a given statement with a specific category membership. (e.g., "Only a woman would say something like this.")

Any assignment of a statement to a speaker can arise through different constellations of these processes. For example, a correct assignment can arise through item and person memory, or in the absence of person memory, through category memory and a lucky guess, as well as through a variety of other process constellations (see Appendix A.2). The model specifies how these different processes work in concert to produce the different kinds of assignments. On the basis of the model, observed assignment frequencies can be used to obtain estimates of the contributions of the different involved processes (Batchelder & Riefer, 1999). The convergent and discriminant validities of the resulting process measures were demonstrated by Klauer and Wegener (1998).

In the present paper, we further extend the paradigm to the case of crossed categories. We focus on two of the identified processes, namely on category memory and reconstructive category guessing. If a discussion statement is recognized as old, but its speaker has been forgotten, perceivers may still remember the speaker's category affiliation and assign the statement to a speaker of the correct category. The strength of category memory as assessed by the model parameter d is a measure of category salience in encoding and interpreting the group discussion (Klauer & Wegener, 1998). Similarly, if an item is judged old, but neither its speaker nor category affiliation can be retrieved, the statement must nevertheless be assigned to someone. Participants are then likely to use the statement content as a cue to make an informed guess about its speaker (Klauer & Wegener, 1998; Klauer et al., 2002; van Knippenberg, van Twuyver, & Pepels, 1994): When stereotypical expectancies have been activated, participants are more inclined to assign the statement to a speaker from the stereotype-consistent category than to a member of an inconsistent category. The strength of this tendency as assessed by the model parameter a provides a measure of reconstructive stereotype application.

For example, in trying to remember who advocated to abandon the use of nuclear energy in a recent parliament debate between members of different political parties, one may still remember the speaker's party affiliation even if there is no memory for the particular party member who made the statement. That is, there may be category memory (i.e., memory for the party affiliation) as assessed by parameter d, implying social categorization during encoding. If not even category membership can be retrieved, one may still be inclined to guess (rightly or wrongly) that it was someone from the pro-environment party on the grounds that the statement is consistent with one's stereotype of that party. That is, category membership is reconstructed through the use of stereotypical associations as assessed by parameter a. It is well documented that stereotype-based guessing biases operate in recognition tasks (Stangor & McMillan, 1992) and in the assignments of the "Who said what?" task (e.g., Klauer et al., 2002; van Knippenberg et al., 1994). Their measurement in that task depends on the use of statements with contents that are stereotypically linked to the employed categories as elaborated below. Like other indirect measures of stereotyping (e.g., priming measures; Wittenbrink, Judd, & Park, 2001a), the guessing-bias measure is unobtrusive because participants are not aware of the fact that their stereotypes are being tapped. Unlike priming or word-fragment-completion measures of stereotyping, the guessing-bias measure involves stereotype application rather than only stereotype activation (Gilbert & Hixon, 1991), because reconstructive category guessing consists of the use of stereotypical expectancies, if and when they are activated, in making an informed guess under uncertainty (but see Lenton, Blair, & Hastie, 2001, for another view). Unlike stereotype measures based on trait ratings, reconstructive category guessing provides a non-evaluative measure of stereotype application in the sense that participants do not intend to evaluate, or to express an evaluative impression, of group members: their goal is to make assignments of statements to speakers that are as accurate as possible.

Structural analysis of category memory and reconstructive category guessing

What are the effect patterns that may arise for category memory and reconstructive category guessing when age and gender categories are crossed? Turning first to category memory: (a) there can be subgroup-level memory in which both the speaker's age category and his or her gender category are remembered, (b) there can be exclusive memory for the age category without memory for the gender category, (c) exclusive memory for the gender category, but not the age category, and (d) neither age nor gender memory. Table 1 shows the probabilities of each event. To characterize the pattern of these probabilities, it is helpful to consider heuristically a simple successive refinement model of category-based impression formation as proposed by Brewer (1988). Brewer suggests that category-based impression formation moves from general and abstract categories to successively narrower categories until a sufficient level of fit between the category and the target person's features is attained: "The search is presumed to continue as an iterative, pattern-matching process, starting at the most general level of categorization and progressing to more specific subtypes, until an adequate fit is achieved." (Brewer, 1988, p. 17). Thus, depending on context factors, among others, one dimension such as gender might be considered first. If an adequate fit is achieved, the process stops implying exclusive gender memory. Sometimes fit will not be adequate, and additional category distinctions such as the age distinction will be considered, leading to subgroup memory. To make this heuristic model stringent, it is necessary to assume that a retrievable memory trace, containing the considered category memberships, is formed with a

Table 1
Joint probabilities of memory (d) for and reconstructive guessing (a) of category membership in crossed categories

Category memory	Age remembered	Gender remembered		
		Yes	No	Sum
	Yes	d(age, gender)	d(age, not gender)	d(age)
	No	d(not age, gender)	d(not age, not gender)	$1 - d$(age)
	Sum	d(gender)	$1 - d$(gender)	1
Reconstructive guessing	Guessing age to be	Guessing gender to be		
		Female	Male	Sum
	Old	a(old, female)	a(old, male)	a(old)
	young	a(young, female)	a(young, male)	$1 - a$(old)
	Sum	a(female)	$1 - a$(female)	1

certain fixed encoding probability at whichever level the process stops. The encoding probability is then reflected in the overall amount of category memory, that is, the sum of the probabilities of subgroup memory and the two kinds of exclusive memory.

In this model, the level at which the process stops determines the relative amount of subgroup memory, whereas the order in which age or gender are initially accessed determines the relative proportion of the two kinds of exclusive category memory. Consequently, we propose to characterize the structure of category memory by two indices: *relative subgroup memory* and *relative category dominance*. Relative subgroup memory is defined as the proportion of cases with subgroup memory among the cases with any kind of category memory, that is, with subgroup memory or (one of the two kinds of) exclusive memory. In the present context, relative category dominance was arbitrarily chosen to reflect relative dominance of gender over age (rather than the other way around), so that it is defined as the proportion of cases with exclusive gender memory among the cases with exclusive gender or exclusive age memory.[4] The two structural indices can vary independently between 0 and 1. Together with the overall amount of category memory they completely characterize the probabilities of Table 1. Thus, their descriptive power does not depend upon the validity of the successive refinement model of impression formation that is used here to point out one possible interpretation of the indices in terms of underlying impression-formation processes. An alternative interpretation in terms of competitive category activation is considered in the General discussion.

Fig. 1 shows the space spanned by relative subgroup memory and relative gender dominance along with a number of specially labeled patterns of category memory. When impression formation always stops as soon as either age or gender is considered, relative subgroup memory will be zero, and the resulting effect patterns are accordingly termed *single categorization* in Fig. 1. When in addition relative gender dominance is one, there is not only no subgroup memory, but also no exclusive age memory, implying that the age category has left no trace at all in category memory, and category memory follows a simple pattern of *gender dominance*: There is memory for gender, but age is completely ignored. Similarly, when relative gender dominance equals zero, there is *age dominance*. In contrast, when gender and age have equal impacts, relative gender dominance equals .5 characterizing *symmetrical patterns*.

As subgroup memory increases, there will be increasing amounts of joint age and gender memory; however, relative gender dominance may still reveal relatively greater impacts of age or gender or a more symmetrical pattern. When relative gender dominance is one, age memory occurs only in combination with gender memory, that is, in the form of subgroup memory; a pattern that we have labeled *hierarchical ordering by gender* in Fig. 1 because age memory is given only if there is also gender memory. As outlined above, one interpretation of this pattern is that the gender dimension is always accessed first in the impression-formation process that sometimes, but not always, proceeds to the subgroup level. Similarly, patterns without exclusive gender memory have been termed *hierarchical ordering by age*. When relative subgroup memory reaches its maximum of one, category memory is either both or none; that is, there is either joint memory for age and gender or no category memory at all. This pattern, termed *merged representation*, arises if the impression-formation process always progresses to the subgroup level. Relative gender dominance is not meaningfully defined for the case of merged representation as there is no exclusive category memory. Note that according to Brewer (1988), categories such as age, gender, or ethnicity are always assumed to be applied in impression formation due to their primordial nature, leading one to expect patterns of merged representation in the present studies.

A final pattern that is somewhat orthogonal to this characterization is the *independence* pattern, corresponding to the additivity pattern observed in measures of intergroup bias. The pattern is given if memory for one category dimension is independent of memory for the other dimension: the probability of remembering subgroup membership is the product of the probabilities of remembering category membership for each of the category distinctions separately. This pattern has no simple interpretation in the above successive refinement model. Although stochastical independence is compatible with the successive refinement model, it should arise only accidentally. The independence pattern is more supportive of a model in which both category dimensions are activated and encoded independently of one another. To conclude the discussion of the category-memory component of the present model, note that an intriguing possibility is that different patterns might govern social categorization for the different subgroups as a function of perceivers' ingroup–outgroup status.

To assess reconstructive category guessing, we employed statements that were subject to multiple stereotypical associations. Statements were presented that were found to be more typical for female speakers and simultaneously more typical of young speakers in situations with single categorizations. Simultaneously, other statements were more typical of male speakers as well as

[4] Relative subgroup memory $= d(\text{age, gender})/(d(\text{age, gender}) + d(\text{not age, gender}) + d(\text{age, not gender}))$; relative gender dominance $= d(\text{not age, gender})/(d(\text{not age, gender}) + d(\text{age, not gender}))$; overall amount of category memory $= d(\text{age, gender}) + d(\text{not age, gender}) + d(\text{age, not gender})$.

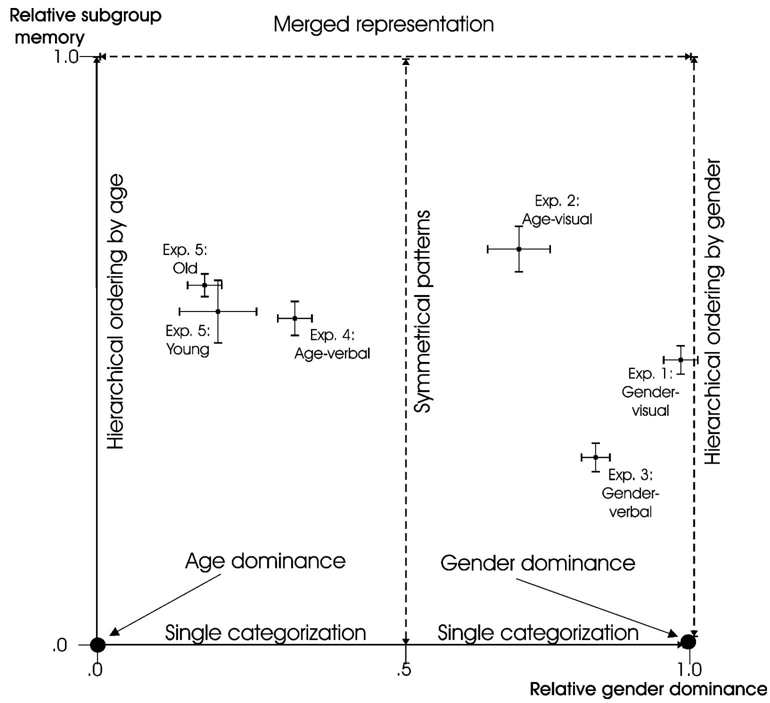

Fig. 1. Relative subgroup memory and relative gender dominance in Experiments 1–5. The error bars mark off one half of the standard error of relative gender dominance left and right of the parameter estimate, and one half of the standard error of relative subgroup memory below and above the parameter estimate. The x-axis is given by relative gender dominance. Relative gender dominance can vary between 0 and 1 and is defined as the proportion of exclusive gender memory relative to the sum of exclusive gender memory and exclusive age memory. The y-axis is given by relative subgroup memory. Relative subgroup memory also varies between 0 and 1 and is defined as the proportion of subgroup memory relative to the total amount of category memory (subgroup memory and the two kinds of exclusive category memory).

of old speakers. Dominance in reconstructive category guessing is given if the stereotypical expectancies associated with one of the dimensions are ignored, while the other set of expectancies is used in reconstructive guessing. Independence is given if both guessing tendencies operate independently and simultaneously. The same models as defined for intergroup bias can be considered for the extent of reconstructive category guessing that is associated with each subgroup as a function of its ingroup–outgroup status. Because the patterns observed for reconstructive category guessing turned out to be very simple in the present studies, we omit an elaborated discussion of the possible effect patterns for this measure for the sake of brevity.

Previous findings

A number of studies have looked at measures of social categorization in situations with crossed categories using the "Who said what?" paradigm (Arcuri, 1982; Biernat & Vescio, 1993; Stangor, Lynch, Duan, & Glass, 1992; van Knippenberg et al., 1994; van Twuyver & van Knippenberg, 1995; van Twuyver & van Knippenberg, 1998; Vescio, Judd, & Kwan, in press). Of these, Arcuri (1982), Biernat and Vescio (1993), van Twuyver and van Knippenberg (1995), van Knippenberg et al. (1994), and Vescio et al. (in press) do not focus on analyzing the structure of dependencies between the crossed category distinctions. However, Biernat and Vescio (1993) and Vescio et al. (in press) report interactions of the two employed category dimensions in the observed confusion frequencies.

Stangor et al. (1992) investigated the effects of crossing ethnicity and gender (Exps. 1–3), and ethnicity and clothing style (Exps. 4 and 5). van Twuyver and van Knippenberg (1998) summarize a number of studies in which gender categories and categories defined by academic status were crossed. Both series of studies rejected the independence model and found evidence for a hierarchical ordering of social categorization in memory. That is, one dimension, usually gender, was superordinate, and there was substantial evidence for exclusive memory for the superordinate dimension, but little evidence for exclusive memory along the subordinate dimension. In one experiment (Exp. 5), Stangor et al. (1992) found a dominance pattern in which ethnicity was the only dimension encoded, and a contextually irrelevant feature, clothing style, was ignored.

There are, however, a number of problems with these results that limit their conclusiveness. First, no provision was made to control for reconstructive category guessing. As argued above, when guessing under uncertainty, participants may be differently inclined to assign statements to the different subgroups, rendering the interpretation of the traditional error-difference measure problematic. This problem may well be of minor importance given that many of the above-mentioned studies used statement contents that were not particularly relevant for the categories in question, so that guessing may not have favored one subgroup over others. More importantly, the traditional analysis in terms of raw confusion frequencies are biased in favor of artefactually detecting dependencies between category dimensions, that is, in favor of producing spurious evidence for categorization at the subgroup level and simultaneously in favor of rejecting exclusive memory for a given dimension. This is due to the fact that the different kinds of confusions are confounded with statement memory in the "Who said what?" paradigm as detailed in Appendix B.

To our knowledge, there are even fewer studies that have directly investigated the use of stereotypical expectancies in situations with crossed categories. Nevertheless, research on subtypes (e.g., Brewer, Dull, & Lui, 1981) demonstrates that distinctive stereotypes can develop for subgroups of broad superordinate categories such as that of old persons. For example, it is plausible that stereotypical expectancies activated for a young man differ from those activated for a young woman, an old man, or an old woman. The question is whether these different reactions reflect effects of the conjunction of social categories rather than a simultaneous, but independent emphasis on different age and gender cues. Indeed, subtyping is assumed to occur in stereotyping if such differences do not simply go back to the joint activation of more than one social category. For example, stereotypical expectancies associated with young women may contain unique beliefs that are not part of broader stereotypes of women or young persons. In terms of reconstructive category guessing, conjunctive effects of age and gender should lead to the rejection of the independence model. On the other hand, Macrae et al. (1995) have argued that in the presence of multiple stereotypes, one stereotype becomes dominant, whereas the other set of stereotypical associations is inhibited, leading one to expect a simple dominance pattern in stereotype application (cf. Bodenhausen & Macrae, 1998; Sinclair & Kunda, 1999).

Finally, the effect patterns of social categorization as well as stereotyping are likely to be sensitive to a host of context variables that enhance the salience of one category dimension over the other. Various measures of social categorization have been found to be affected by priming procedures (Macrae et al., 1995, Exp. 1; van Twuyver & van Knippenberg, 1995), presentational setting (Macrae et al., 1995, Exp. 3), motivational variables (Pendry & Macrae, 1996, Exp. 1; Sinclair & Kunda, 1999; Stangor et al., 1992), and comparative context (van Rijswijk & Ellemers, 2002), among others. These findings suggest that variations in the stimulus context affect the extent to which a category cue triggers social categorization of a target person as well as the extent to which a category cue leads to activation of stereotypic group-related memory contents (Wittenbrink, Judd, & Park, 2001b). As a reasonable first step in validating the new method, we manipulated context relevance (Hewstone, Hantzi, & Johnston, 1991; van Knippenberg et al., 1994) by using different discussion topics in different studies and sought to demonstrate that the method can (a) recover the major findings of previous studies reviewed above and (b) to demonstrate how these findings can be extended and refined through the new experimental and data-analytic strategies proposed here.

Experiments 1–4

In Experiments 1–4, age and gender categories were crossed in order to investigate the pattern of social categorization in memory and of reconstructive category use in the "Who said what?" paradigm. To assess possible context effects, context relevance was varied for the two category distinctions via discussion topic; orthogonally, speakers and their category membership were either identified by means of visual or verbal cues.

Specifically, in Experiments 1 and 3, the topic of the discussion was gender roles in intimate relationships and in society at large, implying context relevance for gender, and to a lesser extent for age. Conversely, in Experiments 2 and 4, the discussion was about the state of affairs at the university, and since old persons were said to be teachers and young persons students, there was context relevance for age, and to a lesser extent for gender in these experiments.

When the speakers' category memberships have to be extracted from photographs as in Experiments 1 and 2, confusions in the "Who said what?" paradigm might be guided by perceptual similarity rather than category memory (Stangor et al., 1992). In addition, visual age cues might differ for the different genders. For example, thinning hair could be more closely associated with old age in men than with old age in women, whereas getting wrinkles might be more typical of older women than of older men, raising the possibility that dependencies in category memory reflect relatively superficial dependencies in the visual age and gender cues. For these reasons, speakers and their category membership were identified by means of verbal labels in Experiments 3 and 4. In what follows, we will refer to Experiments 1–4 as the gender-visual study, the age-visual study, the gender-verbal study, and the age-verbal study, respectively, to emphasize context relevance (for the gender dimension versus for the age dimension) and kind of category cue (visual versus verbal) whenever we feel that these labels increase readability.

In each study, there were twelve discussants, three in each subgroup defined by a combination of age and gender. Each experiment implemented an additional within-participants factor "kind of statement:" For each discussion topic, statements could be of two kinds; for the gender-roles context, there were relatively conservative and relatively progressive statements; for the university context, there were critical and more lenient statements. Each discussant made two statements of each kind so that there was no comparative fit (Oakes, Haslam, & Turner, 1994). In previous experiments with single categorization, it was found that in reconstructive category guessing there was a preference to assign conservative statements about gender roles to men and more progressive ones to women (Gawronski, Ehrenberg, Banse, Zukova, & Klauer, 2003, Exp. 2), and an equally strong tendency to assign conservative statements to old speakers and more progressive ones to young persons (Klauer et al., 2002, balanced condition). Similarly, in single categorization, critical statements about the state of affairs at the university were found to be more strongly associated with men than with women, and vice versa for the more lenient statements (Klauer & Wegener, 1998, Exp. 4). The university statements as originally generated by students from the sampled population were all critical whereas we had to generate corresponding lenient statements ourselves. For this reason, we also assume, although did not separately pretest, that critical statements would be preferentially assigned to students rather than teachers and vice versa for more lenient statements in a situation with single categorization.

Method

Materials. There were two large pools of 192 statements each, one on gender roles (cf. Gawronski et al., 2003, Exp. 2; Klauer et al., 2002, and Wegener, 2000) and one on conditions at the university (cf. Klauer & Wegener, 1998, Exp. 4; Klauer et al., 2002, pilot study). Half of the gender-roles statements expressed a more conservative, the other half a more progressive point of view. Conservative and progressive statements were constructed in comparably worded pairs so that for each statement, there was a parallel one expressing the alternative attitude towards the same or a very similar aspect. For example, the progressive statement "It is only natural that a married man has to iron his shirts himself" finds its counterpart in "It is somewhat strange if a married man has to iron his shirts himself." Analogously, half of the university statements expressed satisfaction with current conditions whereas the other half expressed a critical point of view. Again, statements formed pairs. For example, the parallel version of the critical statement "Most teachers don't take care that course contents match exam requirements" was "Most teachers take care that course contents match exam requirements."

Within both pools, statements were classified into four subtopics consisting of 24 statements each, and each speaker made one statement about each of these subtopics. This was to prevent participants from perceiving discussants as holding contradictory views on the same or a very similar issue. Statements were chosen randomly from these pools and were randomly assigned to speakers for each participant anew under the restrictions that each speaker made one statement on each subtopic, two statements of each kind, and that no two parallel statements were ever shown to the same participant.

For the two studies with visual category cues, six pictures of speakers were chosen for each gender $\times$ age subgroup from a large pool of black and white portrait photographs that were rated to be unambiguous with respect to age category membership by ten judges. The young speakers were perceived to be in their twenties, the old ones in their forties. From the six speakers of each subgroup, three were randomly chosen for each participant anew to be presented in the discussion. For the two studies with verbal category cues, 24 easy to discriminate first and last names were chosen that were rated to be of medium popularity by three judges and that unambiguously identified the named person's gender through the first names. Names were matched in length, and an age indication (ranging from 21 to 26 and from 43 to 48 years, respectively) was randomly assigned to each name for each participant, so that each name could identify both old and young speakers across participants. Six members of each gender $\times$ age subgroup (e.g., Anja Hofmeister, 24 years) were thereby defined and three were randomly chosen for each participant anew to appear in the group discussion.

Participants. 43, 40, 40, and 41 participants took part in the gender-visual study, the age-visual study, the gender-verbal study, and the age-verbal study, respectively. In each study, there were at least 20 male and 20 female participants, all of them students in their twenties with different majors. They either received partial course credit or a small monetary gratification.

Procedure. Experimental sessions were conducted on an individual, computerized basis. Participants read that they were to take part in a study on impression formation, were to observe a group discussion on either gender roles or conditions at the university, and that their task was to form an impression of the group as a whole. Participants of the age-verbal and age-visual study (university topic) were additionally told that the younger discussants were students and the older ones teachers. They then watched the simulated discussion. Each contribution consisted of a written statement along with either a portrait photo in the visual studies or a name and age label identifying the speaker in the verbal studies. Contributions were presented for 6 s with an inter-stimulus interval of 0.5 s. Speakers made their statements in turns so that in a first round, all twelve discussants made their first statement followed by a second round of statements by each speaker and so forth. The order of speakers was randomized in each round anew.

After observing the discussion, participants were shown, in random order, all 48 statements mixed with 48 distracter statements sampled in the same manner as discussion statements. Their task was to decide first whether the statement had been previously presented by clicking on screen fields labeled "old" and "new." If they decided "old," the 12 speakers were presented (either by photographs or by verbal name and age indications), and participants were asked to assign the statement to its speaker by clicking on one of the 12 speakers. After a "new" decision or a person decision, the next statement appeared and was to be classified. Finally, participants were asked to speculate on the purpose of the study, thanked and debriefed.

Data analyses. Preliminary tests determined whether any systematic ingroup–outgroup differences according to participants' gender were present in the assignment data. This was not the case (see Appendix A.3), and the data were accordingly collapsed over participants' gender for further analyses. The data of Experiments 1–4 were analyzed jointly using the multinomial model described in more detail in Appendix A.2 to allow us to compare results between experiments. The model was fitted with different parameters for every experiment. A number of straightforward technical assumptions allowed us to set some parameters equal across conditions and thereby to simplify the model. These restrictions were tested by the initial goodness-of-fit test and are reported in detail in Appendix A.3.

As elaborated in the introduction and Appendices A.1 and A.2, the model provides parameter estimates for the following processes: item detection D, distracter detection D_N, item status guessing b, person memory c, and of central interest here, it extends Klauer and Wegener (1998) model by category memory parameters for the speaker's subgroup d(age, gender), exclusive age memory d(age, not gender), exclusive gender memory d(not age, gender), and three parameters for category guessing as the probabilities to assign a given statement to a particular subgroup under uncertainty: for the subgroup of old women a(old, female); of young women a(young, female); of old men a(old, male). The category guessing parameter for young men is given by one minus these probabilities, and the probability of neither age nor gender memory by one minus the above category memory parameters. All of these parameters, except that for guessing the item status as old versus new, were allowed to vary between experiments, and the parameters for category guessing were additionally allowed to vary as a function of kind of statement. The parameters of the resulting baseline model are estimated using the maximum-likelihood method as explained by Batchelder and Riefer (1999). Based on these parameter estimates, model predictions are then computed for the observed assignment frequencies, and a chi-squared distributed goodness-of-fit statistic assesses how closely these predictions match the observed assignment frequencies. Non-significant values of the goodness-of-fit statistic indicate that the deviations of the model predictions from the data are not significant, that is, that the model fits the data. The baseline model has 144 degrees of freedom, and it accounts for the observed assignment frequencies very well as indicated by the chi-squared distributed goodness-of-fit value: $\chi^2(144) = 143.49$, $p = .50$.

Departing from a baseline model, focused hypotheses tests can be conducted in the multinomial-modeling framework by comparing the baseline model with a more restricted model (Batchelder & Riefer, 1999). For example, to test for stochastical independence of the two dimensions of categorization in category memory, the three model parameters for category memory d(age, gender), d(not age, gender), and d(age, not gender) were replaced by appropriate products of two parameters for group-level category memory d(age) and d(gender), and thus $d(\text{age, gender}) = d(\text{age})\ d(\text{gender}), d(\text{not age, gender}) = (1 - d(\text{age}))\ d(\text{gender}), d(\text{age, not gender}) = d(\text{age})(1 - d(\text{gender}))$ (implying necessarily that $d(\text{not age, not gender}) = (1 - d(\text{age}))(1 - d(\text{gender})))$. This defines a more restricted model, in which category memory is restricted to follow a pattern of stochastical independence. A chi-squared distributed log-likelihood ratio statistic can then be computed to test whether the restriction leads to a significant deterioration of fit relative to the baseline model. A significant outcome of this test indicates that the specific restriction introduced in the restricted model (i.e., in the present example, stochastical independence in category memory) is not compatible with the data (see also appendix in Klauer & Wegener (1998), for further explanations of the logic of hypothesis testing in multinomial modeling).

Results

First we conducted tests for stochastical independence of the two dimensions of categorization in category memory and in reconstructive guessing. For category memory, there are substantial dependencies across the four experiments, $\chi^2(4) = 76.08, p < .001$, and for each experiment individually (for Exps. 1–4, respectively, $\chi^2(1) = 16.25, \chi^2(1) = 34.50, \chi^2(1) = 4.97, \chi^2(1) = 22.70$; in each case $p < .05$). In each experiment, the amount of subgroup memory was larger than expected under the independence model. In contrast, for reconstructive category guessing, independence can be maintained for all studies and both kinds of statements: $\chi^2(8) = 10.36, p = .24$ (for Exps. 1–4, respectively, $\chi^2(2) = 2.76, p = .25$; $\chi^2(2) = 2.03, p = .36$; $\chi^2(2) = 4.78$, $p = .09$; $\chi^2(2) = 0.78, p = .68$). Thus, reconstructive category guessing can be described by the separate and independent guessing tendencies for gender and for age for each kind of statement and experiment, and subsequent analyses were based on a model that incorporated this simplification. That is, the subgroup-level parameters were replaced by two parameters, one for gender as the probability to assign a statement to a woman a(female) rather than a man $(1 - a(\text{female}))$ under uncertainty, and one for age as the tendency to assign a statement to an old person a(old) rather than a young person $(1 - a(\text{old}))$, for each kind of statement and experiment.[5] The simplified model has 152 degrees of freedom and fits the data very well ($\chi^2(152) = 153.85, p = .44$). The parameter estimates and their confidence intervals are shown in Table 2.

Category memory. Based on the category memory parameters, we computed the indices for relative gender dominance and relative subgroup memory, and their values (and standard errors) are displayed in Fig. 1 for each study. Relative gender dominance significantly departed in each study from the value .5 that characterizes symmetrical patterns (for Exps. 1–4, respectively, $\chi^2(1) = 86.92, \chi^2(1) = 3.94, \chi^2(1) = 51.44, \chi^2(1) = 6.17$; in each case $p < .05$). As can be seen in Fig. 1, in the gender-visual study, the gender-verbal study, and in the age-visual study, the impact of gender outweighed that of age. In fact, relative gender dominance was not significantly different from one in the gender-visual study ($\chi^2(1) = 0.01, p = .92$) and in the age-visual study ($\chi^2(1) = 2.08, p = .15$), indicating that hierarchical ordering with gender as the superordinate category describes the pattern of categorization in these studies. Relative gender dominance was significantly smaller than one in the gender-verbal study ($\chi^2(1) = 4.10$, $p < .05$), but as can be seen in Fig. 1, the pattern is still best approximated by hierarchical ordering by gender. As already shown above, relative gender dominance also deviated from .5 in the age-verbal study. Relative gender dominance was however, significantly smaller than .5 in this case, implying greater weight for age than for gender, and it was also significantly larger than zero ($\chi^2(1) = 8.37, p < .05$), that is, a significant amount of exclusive gender memory remained.

Relative subgroup memory was significantly larger than zero in each study as already implied by the tests for stochastical independence. Relative subgroup memory was also significantly smaller than one in each study (for Exps. 1–4, respectively, $\chi^2(1) = 123.56$, $\chi^2(1) = 11.79, \chi^2(1) = 97.41, \chi^2(1) = 29.44$; in each case $p < .05$). Thus, the patterns were set apart from both the case labeled single categorization as well as that labeled merged representation in Fig. 1.

Then we tested for effects of discussion topic and category-cue modality. With respect to relative gender dominance, pairwise tests between all studies summarized in Table 3 revealed, not surprisingly, that relative gender dominance was significantly smaller in the age-verbal study than in any of the other three studies. In fact, relative gender dominance could be set equal across the gender-visual study, the gender-verbal study, and the age-visual study ($\chi^2(2) = 3.64, p = .16$), and the joint parameter value was not significantly different from one ($\chi^2(1) = 2.57, p = .11$), underlining that the effect patterns are best characterized by hierarchical ordering by gender in these studies.

Pairwise comparisons of studies were also performed for relative subgroup memory, and their results are also shown in Table 3. It can be seen that relative subgroup memory was significantly larger in the age-visual and the age-verbal study (university context) than in studies with gender context when category-cue modality (visual versus verbal) was kept constant. Conversely, relative subgroup memory was not significantly affected by category-cue modality within these contexts. It was in fact possible to set equal relative subgroup memory across cue modality within each context ($\chi^2(2) = 4.95, p = .08$), and relative subgroup memory was significantly larger in the university context (index value = .61) than in the gender-roles context: index value = .43; $\chi^2(1) = 7.29$, $p < .01$.

Reconstructive category guessing. Fig. 2 shows the differences in the age and gender guessing tendencies between kinds of statements as a function of discussion topic and verbal versus visual category cues. For the university context, these were the differences between the guessing parameters for lenient minus critical statements; for the gender-roles context, the differences between the progressive and the more conservative statements. The differences were signed so that positive values indicate the use of stereotypical expectancies as found in situations with single categorization, that is, a consistency bias in guessing. For example, in the gender context, a positive difference for guessing age categories means that conservative statements were more likely attributed to old speakers than progressive statements under uncertainty (and vice versa for young speakers); a positive difference for guessing gender categories means that progressive statements were more likely attributed to female speakers than conservative statements under uncertainty (and vice versa for male speakers).

We tested whether reconstructive category guessing is best described by dominance (one category distinction is

[5] Thus, subgroup-level parameters were replaced by appropriate products of the age and gender parameters or their complements: $a(\text{old, female}) = a(\text{old})\ a(\text{female})$; $a(\text{young, female}) = (1 - a(\text{old}))\ a(\text{female})$; $a(\text{old, male}) = a(\text{old})(1 - a(\text{female}))$; $a(\text{young, male}) = (1 - a(\text{old}))(1 - a(\text{female}))$.

Table 2
Parameter estimates (PE) and 90% confidence intervals (CI) for Experiments 1–4

Process (parameter)	Category cue	Context relevance	Statement	PE	CI
Item recognition (D)	Visual	Gender	=	.76	.74–.78
		Age	=	.71	.68–.73
	Verbal	Gender	=	.70	.68–.72
		Age	=	.68	.66–.70
Distracter detection (D_N)	Visual	Gender	=	.76[a]	.74–.78
		Age	=	.56	.48–.65
	Verbal	Gender	=	.64	.57–.71
		Age	=	.56	.47–.64
Item status guessing (b)	=	=	=	.19	.16–.22
Person memory (c)	Visual	Gender	=	.38	.35–.41
		Age	=	.20	.17–.22
	Verbal	Gender	=	.09	.07–.12
		Age	=	.07	.04–.10
Subgroup memory (d(age, gender))	Visual	Gender	=	.35	.31–.39
		Age	=	.36	.33–.39
	Verbal	Gender	=	.19	.00–.52
		Age	=	.33	.29–.38
Exclusive gender memory (d(not age, gender))	Visual	Gender	=	.35	.30–.42
		Age	=	.12	.07–.17
	Verbal	Gender	=	.33	.09–.58
		Age	=	.09	.04–.14
Exclusive age memory (d(age, not gender))	Visual	Gender	=	.00	.00–.01
		Age	=	.05	.01–.09
	Verbal	Gender	=	.06	.00–.54
		Age	=	.17	.12–.23
Category guessing gender to be female (a(female))	Visual	Gender	Conservative	.46	.40–.53
			Progressive	.45	.39–.52
		Age	Critical	.49	.45–.54
			Lenient	.51	.44–.55
	Verbal	Gender	Conservative	.42	.37–.47
			Progressive	.58	.53–.62
		Age	Critical	.49	.44–.52
			Lenient	.46	.42–.50
Category guessing age to be old (a(old))	Visual	Gender	Conservative	.54	.50–.65
			Progressive	.58	.53–.62
		Age	Critical	.43	.38–.47
			Lenient	.55	.50–.59
	Verbal	Gender	Conservative	.53	.49–.56
			Progressive	.49	.45–.52
		Age	Critical	.36	.32–.40
			Lenient	.56	.52–.61

Note. Equality signs indicate that parameters were restricted to be equal over the respective conditions.
[a] Set equal to D(visual, gender).

dominant at the expense of the other which is ignored) or the independence model (both category dimensions exert independent, but significant effects). As can be seen in Fig. 2, in the university context, lenient statements were more likely attributed to old speakers than critical statements (and vice versa for young speakers). That is, stereotypical expectancies associated with the age distinction were used in guessing in the university context. This guessing bias was significant when tested jointly for the age-visual study and the age-verbal study ($\chi^2(2) = 38.94, p < .05$) as well as in individual tests ($\chi^2(1) = 10.55, p < .05$ for the age-visual study, $\chi^2(1) = 28.39, p < .05$ for the age-verbal study). Conversely, there was no evidence for reconstructive category guessing by means of gender-based expectancies in the university context ($\chi^2(2) = 2.17, p = .34$), that is, lenient and critical statements were attributed to female rather than male speakers with equal likelihood under

Table 3
Pairwise tests for equality of relative gender dominance and relative subgroup memory

Index	Study A	Study B	$\chi^2(1)$	p
Relative gender dominance	Gender-visual	Age-visual	1.90	.17
		Gender-verbal	2.26	.13
		Age-verbal	29.43	<.01
	Age-visual	Gender-verbal	0.64	.42
		Age-verbal	8.57	<.01
	Gender-verbal	Age-verbal	27.97	<.01
Relative subgroup memory	Gender-visual	Age-visual	3.90	<.05
		Gender-verbal	3.79	.052
		Age-verbal	0.68	.21
	Age-visual	Gender-verbal	11.47	<.01
		Age-verbal	1.16	.28
	Gender-verbal	Age-verbal	6.63	.01

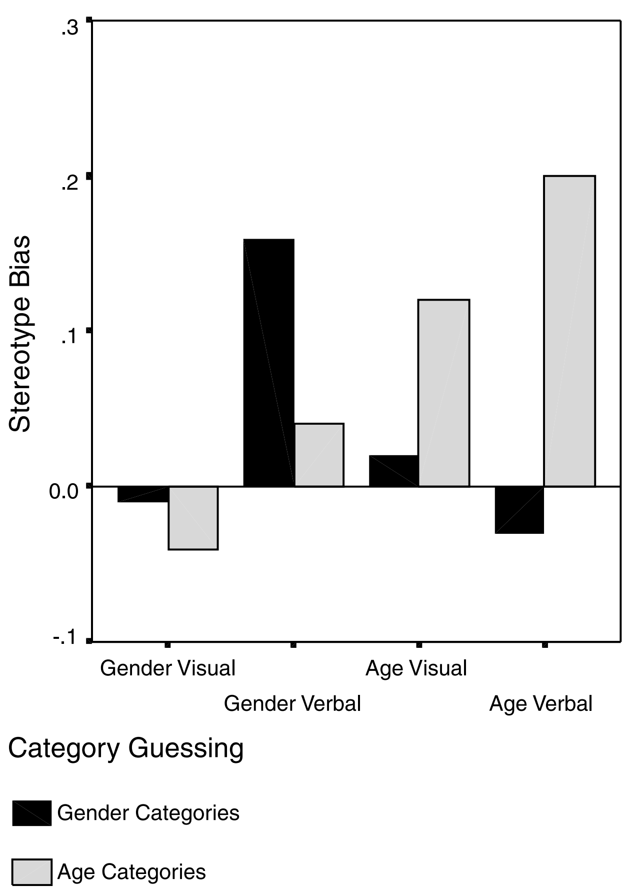

Fig. 2. Amount of stereotype bias in reconstructive age guessing (grey bars) and reconstructive gender guessing (black bars) as a function of discussion context and category presentation format in Experiments 1–4. Positive values indicate stereotype-consistent assignment tendencies.

uncertainty. Thus, the dominance model best describes category guessing in the university context. The context-relevant distinction, age in this case, dominates guessing while the other distinction, gender, is ignored.

The dominance pattern also best describes category guessing in the gender-roles context. As can be seen in Fig. 2, there is now evidence for gender-based reconstructive guessing, progressive statements being assigned more readily to women than conservative statements and vice versa for men. The effect, although significant jointly for the gender-visual and the gender-verbal study ($\chi^2(2) = 14.59, p < .05$), was however, restricted to the gender-verbal study ($\chi^2(1) = 14.55, p < .05$), and was absent when category information was given by photos ($\chi^2(1) = 0.04, p = .84$). Conversely, there was no effect of the age distinction in the studies with gender-roles context ($\chi^2(2) = 0.35, p = .84$), that is, progressive and conservative statements were assigned to old rather than young speakers with equal likelihood under uncertainty. Taken together, there is again evidence for a simple dominance pattern—now with gender taking the superordinate role in the gender-roles context.

Discussion

The results can be summarized succinctly. Hierarchical ordering by gender best described the effect patterns for category salience in memory, and the superordinate role of gender was reversed only in the age-verbal study, in which the age distinction had the greatest impact. In contrast, a simple dominance model governed stereotype application. Like categorization, stereotype application was sensitive to context relevance in that gender and age switched their hierarchical positions depending on which category distinction was more relevant for the discussion topic at hand. An interesting and unforeseen effect of context relevance was that relative subgroup memory was larger in the university context than in the gender-roles context suggesting that participants more frequently found it necessary to access speakers' subgroup membership in the former context. Thus, context relevance can also affect the depth of categorization over and above the relative weight of the two category dimensions.

Overall, similar patterns of effects emerged regardless of whether category cues were visual or verbal, ruling out a perceptual origin of the observed confusions and dependencies and locating them at a more conceptual level of person and category representation in memory. A notable exception to this rule was found in the age-visual study, in which category salience in memory followed a hierarchical ordering pattern with gender rather than age as the superordinate category although age was the context-relevant category. Before discussing the theoretical implications of these findings, we made one more attempt to break the frequently observed priority of gender in social categorization (e.g., Stangor et al., 1992; van Twuyver & van Knippenberg, 1998) over other category distinctions such as age on the basis of visual cues. It would be very interesting indeed if visual gender and age cues always trigger patterns of social categorization in which gender is superordinate independently of variations in stimulus context.

Experiment 5

In Experiment 5, the discussion was again about the state of affairs at the university, and participants were told that the young speakers were first term students (instead of just students as in the previous age-visual study and the age-verbal study) and that the old speakers were professors (instead of just university teachers). In addition, the age categories were more strongly differentiated and individually more salient by presenting pictures of even younger and older male and female speakers.

Method

The procedures closely followed those of the age-visual study (Exp. 2) unless where otherwise mentioned.

Materials. For each gender $\times$ age subgroup, three new black and white portrait photos were selected. As determined by a pretest ($N = 15$), the young persons were estimated to be around 20 years old ($M = 20.33$, $SD = 1.96$) and the old persons were perceived to be in their early seventies ($M = 72.55$, $SD = 4.22$).

Participants. Participants were 40 students in their twenties (20 male and 20 female) with different majors. They received partial course credit or a small monetary gratification for their participation.

Data analyses. Preliminary tests determined whether any systematic ingroup–outgroup differences according to participants' gender were present in the assignment data. This was not the case (see Appendix A.4), and the data were accordingly collapsed over participants' gender for further analyses. The baseline model employed for Exps. 1–4 was fitted to the data set from Experiment 5. The model fit was insufficient, $\chi^2(36) = 62.93, p < .01$. We hypothesized that the more extreme age categories and their members might be differentiably memorizable and thus allowed the person memory and category memory parameters to vary as a function of age. The resulting model achieved a satisfactory goodness of fit, $\chi^2(32) = 40.90, p = .14$, and it was used as baseline model.

Results and discussion

First, we tested for independence of the two dimensions of categorization in category memory and reconstructive category guessing. In category memory, independence can be maintained for old speakers, $\chi^2(1) = 0.93, p = .33$, but must be rejected for young speakers: $\chi^2(1) = 4.78, p < .05$. For reconstructive category guessing, independence can be maintained as in the previous experiments: $\chi^2(2) = 2.71, p = .26$.

Thus, reconstructive category guessing can again be described by the separate and independent tendencies for gender and for age, and the subgroup-level parameters were replaced by two parameters, one for gender as the probability to assign a statement to a woman a(female) rather than a man $(1 - a(\text{female}))$ under uncertainty, and one for age as the tendency to assign a statement to an old person a(old) rather than a young person $(1 - a(\text{old}))$, for each kind of statement.[5] The simplified model fits the data satisfactorily: $\chi^2(34) = 43.61, p = .13$. The parameter estimates and their confidence intervals are shown in Table 4.

Category memory. Fig. 1 shows the structural indices characterizing the patterns of category memory for young and old speakers. Both relative gender dominance ($\chi^2(1) = 0.01, p = .92$) and relative subgroup memory ($\chi^2(1) = 0.08, p = .77$) did not differ significantly between young and old speakers. Nevertheless, the overall amount of category memory (subgroup memory plus exclusive category memory) differed substantially between young and old speakers: $\chi^2(1) = 8.05, p < .01$, category memory for old speakers being superior to that for young speakers. The joint parameter for relative gender dominance was .19, and it was significantly smaller than .5 ($\chi^2(1) = 32.32, p < .01$.), that is, there was a substantially greater impact of age than gender in category memory. Yet, relative gender dominance was also significantly larger than zero ($\chi^2(1) = 5.47$, $p < .05$), so that a significant amount of exclusive gender memory remained and the patterns were only approximated by hierarchical ordering by age. The joint parameter for relative subgroup memory was .61 and thus, the patterns were distinct both from patterns with single categorization (relative subgroup memory $= 0$, $\chi^2(1) = 176.93, p < .01$) and from patterns with merged representation (relative subgroup memory $= 1$, $\chi^2(1) = 80.50$, $p < .01$).

Reconstructive category guessing. As can be seen in Table 4, lenient statements were more likely attributed to old speakers than critical statements (and vice versa

Table 4
Parameter estimates (PE) and 90% confidence intervals (CI) for Experiment 5

Process (parameter)	Category membership	Statement	PE	CI
Item recognition (D)	=	=	.67	.65–.69
Distracter detection (D_N)[a]	=	=	.67	.65–.69
Item guessing (b)	=	=	.24	.21–.26
Person memory (c)	Old	=	.21	.16–.25
	Young	=	.32	.27–.36
Subgroup memory (d(age, gender))	Old	=	.56	.54–.59
	Young	=	.37	.00–.85
Exclusive gender memory (d(not age, gender))	Old	=	.06	.01–.12
	Young	=	.06	.00–.32
Exclusive age memory (d(age, gender))	Old	=	.29	.17–.41
	Young	=	.22	.00–1.00
Category guessing gender to be female (a(female))	=	Critical	.49	.44–.54
	=	Lenient	.47	.42–.52
Category guessing age to be old (a(old))	=	Critical	.22	.16–.28
	=	Lenient	.49	.41–.57

Note. Equality signs indicate that parameters were restricted to be equal over the respective conditions.
[a] Set equal to D.

for young speakers), that is, stereotypical expectancies associated with the age distinction were used in reconstructive category guessing: The guessing bias was significant ($\chi^2(1) = 32.47, p < .01$). Conversely, there was no evidence for reconstructive category guessing by means of gender-based expectancies ($\chi^2(1) = 0.26$, $p = .61$), that is, lenient and critical statements were assigned to female rather than male speakers with equal likelihood under uncertainty. Thus, dominance of the context-relevant stereotype again best describes category guessing in Experiment 5.

Discussion. The results of Experiment 5 can again be summarized succinctly. Category memory for age was much stronger relative to gender memory than in the previous age-visual study (Exp. 2). Age now took the role of the superordinate dimension in category memory, reflecting the use of more strongly differentiated age categories in this study. Category guessing was again best described by dominance: Stereotypical expectancies associated with gender categories as expressed in situations with single categorization were overridden and ignored in favor of the contextually more relevant distinction in terms of age and academic status.

General discussion

The goal of the present work was to introduce and demonstrate the use of a new method for investigating the structure of social categorization and stereotype application. In five experiments, we employed the "Who said what?" paradigm (Taylor et al., 1978) with crossed categories. Following prior work (Hewstone et al., 1991; van Knippenberg et al., 1994), context relevance was manipulated via discussion topic. Models for possible effect patterns in social categorization and stereotype application were formulated and tested.

The new method was first of all able to recover the major findings of previous studies of crossed categorization. Like previous studies (Stangor et al., 1992; Bernat & Vescio, 1993; van Twuyver & van Knippenberg, 1998; Vescio et al., in press), the present studies revealed substantial evidence for social categorization at the subgroup level: The independence model had to be rejected in five out of six tested cases, and in each of these, there was more subgroup-level categorization than expected under independence. The structure of category memory was then further characterized by means of two indices: relative subgroup memory and relative gender dominance. This approach allowed us to formulate effect patterns analogous to those investigated for measures of intergroup bias and to test explicitly whether the observed patterns conformed to or deviated from these possibilities. In the gender-visual study, the age-visual study, and the gender-verbal study, the observed pattern of category memory was best described by hierarchical ordering by gender: In all experiments, there were substantial amounts of gender and age memory. However, age memory came mostly in the form of subgroup memory in these studies, and there was little exclusive age memory. In contrast, in the age-verbal study and in Experiment 5 (visual category cues, context relevance for age), the age dimension had a greater impact than the gender dimension, and the effect patterns were best approximated by hierarchical ordering by age.

These effects of context relevance, although partly based on comparisons between experiments, parallel and refine those observed by van Knippenberg et al. (1994) for relative gender dominance. Relative gender dominance was significantly larger than .5 in studies with a gender-related discussion topic, indicating a greater impact of gender than age, and significantly smaller than .5 in two of three studies with an age-related discussion topic, indicating a greater impact of age than gender, and the differences between studies were also significant. With respect to relative subgroup memory, an interesting and unforeseen effect of discussion topic was that relative subgroup memory was significantly larger in studies with university context than in studies with gender-roles context.

The present studies were the first to assess social categorization and stereotyping for crossed categories in the same experimental context. Although there was strong evidence for social categorization at the subgroup level, stereotype application was governed by simple dominance patterns in which the context-relevant dimension was used and the other dimension was ignored. Generally speaking, the present studies thereby add to a growing body of evidence suggesting that different effect patterns can arise for different dependent measures in crossed categorization (e.g., Brewer et al., 1987; Vanbeselaere, 1991) although on the basis of more cognitive than evaluative dependent measures.

More specifically, the qualitatively different effect patterns underline the importance of distinguishing between social categorization and stereotype activation and application (Lepore & Brown, 1997). Throughout experiments, both category distinctions were frequently extracted, encoded, and used in retrieval as demonstrated by significant relative subgroup memory. Still, stereotypical expectancies associated with the contextually less relevant distinction were completely ignored in stereotype application. This agrees well with current models of stereotype inhibition according to which one set of stereotypical expectancies will come to dominate and suppress alternative sets in situations with multiple stereotypes. As Bodenhausen and Macrae (1998, p. 11) put it "many are called but few are chosen."

The structural indices proposed here for characterizing category memory lend themselves to a simple interpretation in terms of a successive refinement model of category-based impression formation (Brewer, 1988), in which impression formation begins with a broad category such as age or gender and moves to specific subgroups only under certain conditions such as when a sufficient level of fit is not achieved at the first stage. In this framework, relative subgroup memory measures the average depth attained by this process of successive refinement, whereas relative category dominance is determined by which category dimension is preferentially accessed first. Greater relative subgroup memory in the university context than in the gender-roles context means in this interpretation that participants more often found that a broad category such as age or gender did by itself produce sufficient fit in evaluating the discussants and their contributions in the gender-roles context. According to the results for relative gender dominance, that broad category was given by gender. Conversely, age and gender category memberships more often had to be considered jointly to achieve a sufficient level of fit in the university context.

Although Brewer's (1988) theory of nested category structure was used as a framework for the category-memory component of the present model, it should be noted that Brewer, like other theorists, explicitly postulated that age, gender, and race are dimensions on which social stimuli are automatically and simultaneously classified prior to subcategorization. According to her model, categorization by age and gender should have produced patterns of merged representation rather than being hierarchically organized. Some evidence for this joint categorization is provided by Brewer and Lui (1989) who used a free sorting task to access categorization of facial photos. In their data, 93% of co-sortings were of photos that shared both sex and age categorization. All subclusterings were within these age-gender partitions. In contrast, the present work along with previous studies using the "Who said what?" task and categories such as gender, age, or ethnicity reveals a relatively strong dominance of gender over age categorization that is furthermore modulated by context relevance. Although there were substantial amounts of subgroup memory in all our studies, there were also frequently substantial amounts of exclusive gender memory or exlusive age memory, implying that at least sometimes stimuli were categorized by gender irrespective of age or by age irrespective of gender. This suggests: (a) that even primordial dimensions such as age and gender are not always obligatorily and simultaneously accessed and (b) that memory-based measures may be more sensitive in uncovering the finer structure of social categorization with strong categories than sorting tasks. One possibility is that participants in sorting tasks simply feel that their classifications should as a minimum reflect visually prominent distinctions between facial photos as provided by gender and age cues. In fact, instructions to sort according to similarity in physical features and instructions to sort according to similarity in character or personality type produced very similar clusterings in the Brewer and Lui (1989) data.

As already explained, hierarchical ordering in social categorization as observed here and in previous studies flows naturally from a successive-refinement model of social categorization: Contextual priming leads perceivers to access the primed category first (Wyer & Srull, 1989), and only if that category membership is insufficient to make sense of the target person and his or her behavior, they will move on to the subgroup level. Taken together, hierarchical ordering emerges, in which there can be exclusive memory for the dominant category as well as subgroup-level memory, but no exclusive memory for the non-dominant category.

Another explanation can be based on the assumption of competitive category activation. In a situation with multiple category memberships, Macrae et al. (1995) found that the stereotype associated with a primed category was activated, whereas the stereotype of the non-primed category was inhibited. More precisely, relative to a control-group condition with no priming at all, there was more stereotype activation for the primed category than in the control condition, whereas there was less stereotype activation for the non-primed category relative to the control condition. As pointed out by van Knippenberg and Dijksterhuis (2000), it is possible that this competitive activation of a primed stereotype at the expense of a non-primed stereotype already operates at the level of category activation. That is, following a rapid identification of the category memberships of a target person, more activation might quickly accrue for a primed category membership at the expense of activation for a non-primed category even if the non-primed category should be chronically more accessible than the primed category. In the present studies, the activation level of the contextually primed categories would thereby quickly come to equal and eventually exceed that of the contextually irrelevant (non-primed) categories. Assuming that category memory depends on the strength of category activation in perceiving the original episode, it follows that only exclusive memory for the primed category as well as subgroup-level memory is consistent with the resulting pattern of activation. In particular, exclusive memory for the non-primed category cannot arise because it would require substantially greater activation of the non-primed than the primed category membership in this interpretation. The level of subgroup memory, on the other hand, depends upon the extent to which category activation of the non-primed category is depressed. Taken together, hierarchical ordering in category memory can also be explained by a process of competitive category activation.

It is interesting to note that in subsequent stereotype activation and use, competition between multiple categories often leads to patterns of category dominance (Bodenhausen & Macrae, 1998; cf. Sinclair & Kunda, 1999), and category dominance also emerged in the present studies in our category-guessing measure of stereotype use. This means that the processes of excitation and inhibition underlying competitive category activation continue to operate throughout stereotype activation and possibly stereotype use and eventually effectively suppress the non-dominant stereotype rather than merely reducing its activation to levels below that of the dominant stereotype (and below control conditions without priming). Alternatively, it can be speculated that competitive activation becomes a more prominent principle in stereotype than in category selection for functional reasons: The costs of having to deal with two different sets of activated stereotypic beliefs may be much larger than the costs of being aware of two category memberships. Thus, participants can be aware of and later remember multiple categories (although they would not remember the subordinate category without also remembering the superordinate one according to hierarchical ordering), but for stereotype use, only one set of stereotypic beliefs would eventually be selected in a "winner-takes-all" fashion, namely those associated with the dominant category.

To discriminate between the two possibilities to account for hierarchical ordering in category memory, that is, by successive refinement versus by competitive category activation, an interesting use of the present method might be to manipulate the time that is available to participants for encoding each speaker and his or her statement. Given a process of successive refinement, if exposure duration is shortened, the index for subgroup memory should decrease while the index for category dominance should remain unchanged. In contrast, under a model with competitive category activation, depending upon the exact time course of excitation and inhibition processes, the index for category dominance should become less polarized as the processes of excitation and inhibition are given less opportunity to pull apart the activation levels of competing categories as exposure duration is shortened.

Over and above providing detailed structural analyses of categorization and stereotype application, the new method has additional advantages that were not, however, the focus of the present paper. For example, the method allows one to assess the amount of individuation through the person memory parameters independently of the amount of category memory and reconstructive category guessing as explained in Appendix A (cf. Klauer & Wegener, 1998). Perhaps most importantly, the method makes explicit the assumptions upon which it is based and subjects them to an empirical test in the form of the model's goodness-of-fit test whereas the assumptions underlying traditional indices and approaches that were proposed in this context usually remain implicit and untested. As outlined in the introduction and further discussed in Appendices A.2 and B, the new method also controls for a number of methodological problems that otherwise would have provided uninteresting alternative explanations of the findings.

The present studies aimed at illustrating the usefulness of the new method for investigating the structure of social categorization and non-evaluative stereotyping, but they are limited in linking these findings to the differentiated findings on evaluative measures of intergroup bias reviewed in the introduction. Since we did not obtain measures of intergroup bias, it is not possible to test hypotheses about the link between categorization and intergroup bias. However, an intriguing finding is the pervasive absence of ingroup–outgroup differences in social categorization and non-evaluative stereotype application (cf. Appendices A.3 and A.4). That is, the same patterns governed social categorization and reconstructive category guessing irrespective of ingroup–outgroup status of participants and perceived group members. In line with early accounts in terms of category differentiation (Deschamps & Doise, 1978), this suggests that social categorization determines who is seen as belonging together or as belonging to different subgroups, and that it is only in evaluating the perceived partitions that one's own membership in, and relationship to, the perceived subgroups comes into play.

Recent work on intergroup bias has begun to identify the conditions under which different patterns occur. As reviewed by Crisp, Ensari, Hewstone, and Miller (in press), a number of factors cause patterns of intergroup bias that deviate from the frequently observed additivity pattern. For example, increasing the salience of one of the involved category dimensions often leads to category dominance. An interesting idea proposed by Crisp et al. is that the influence of such factors can be mediated either by effects on social categorization or through an affective-priming route. The social-categorization route determines who is seen as belonging together or as belonging to different partitions of the total group. The affective-priming route determines how these partitions are evaluated. Different moderators of intergroup bias are assumed to exert their influences through either one or the other or both of these routes. For example, manipulating the relative salience of category dimensions through priming should only affect the social-categorization part of this model, whereas manipulating perceivers' mood state should primarily affect the affective-priming part of the model, or possibly both parts. Tests of these and similar theoretical assumptions will crucially depend upon the availability of appropriate, and equally sensitive, measures of social categorization and intergroup bias. An interesting application of the method proposed here can fulfill this purpose by providing unobtrusive measures of intergroup bias and social categorization in one and the same experimental paradigm. For this purpose, instead of statements of different kinds, positive and negative bits of information about the individual group members such as positive and negative behaviors can be presented. Reconstructive category guessing in the later assignment test then measures differential tendencies to assign positive and negative information to the subgroups under

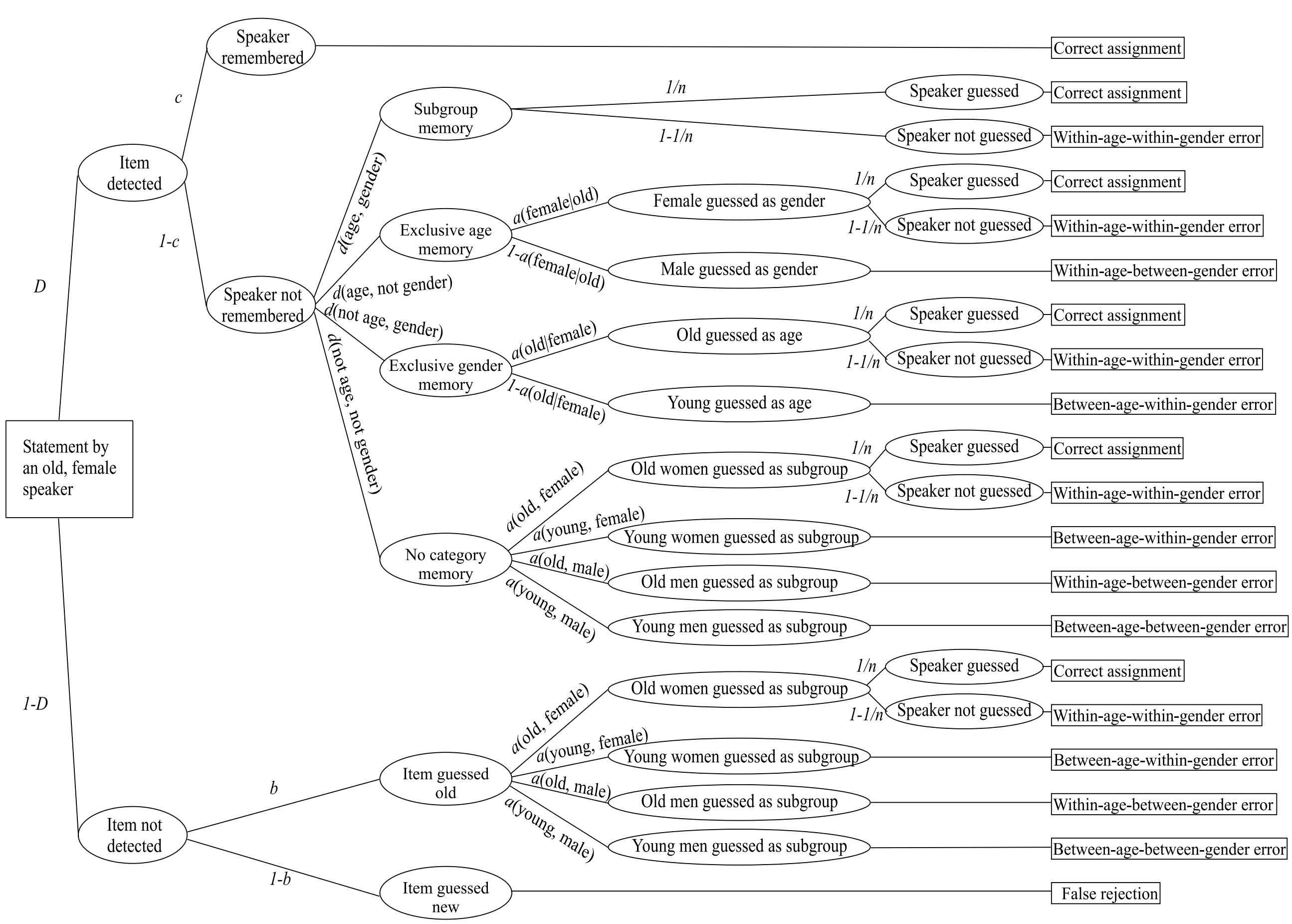

Fig. 3. The multinomial model of crossed social categorization in the modified "Who said what?" paradigm; the partial tree for statements made by a member from the subcategory of old women is shown as explained in Appendix A.2. Response categories are shown in rectangles to the right; they are defined in Table of Appendix A.2. Correct assignments correspond to response Category 1 in Table of Appendix A.2; within-age–within-gender errors to Category 2; between-age–within gender errors to Category 3; within-age–between-gender to Category 4; between-age–between-gender to Category 5; false rejections to Category 6.

uncertainty and was found to be a valid and sensitive measure of group evaluations in an illusory-correlation context (Klauer & Meiser, 2000). Clearly, the new method thereby also has the potential to open an exciting new window on the interplay of evaluative processes and those of social categorization.

Appendix A. Structure, assumptions, and tests of the multinomial model of crossed categorization

A.1. Process parameters

The model for crossed categorization proposed here is a straightforward extension of the source monitoring model validated by Klauer and Wegener (1998). The extensions basically consist in incorporating additional memory and guessing parameters for a second category dimension as illustrated in Fig. 3. Category memory parameters are d(age, gender) for subgroup memory (i.e., for joint memory of age and gender), d(age, not gender) for exclusive age memory without gender memory, and d(not age, gender) for exclusive gender memory. For the sake of clarity, we also use the parameter d(not age, not gender) for the absence of category memory in Fig. 3, but it is not an independent parameter as it is given by one minus the other category memory parameters. An analogous extension is required for reconstructive category guessing to account for possible category guessing at the subgroup level. The category guessing parameters are a(old, female) for the probability to assign a statement to an old woman under uncertainty, a(young, female), and a(old, male). For the sake of clarity, we also use, in Fig. 3, the parameter a(young, male) for the fourth subgroup, but it is not an independent parameter as it is given by one minus the other category guessing parameters. Furthermore, we have to account for the case that there is exclusive memory for age without gender memory, so that gender has to be guessed conditionally on a specific age category membership. For example, a(female | old) denotes the probability to choose a female rather than a male speaker among the old persons under uncertainty in Fig. 3. Similarly, when there is exclusive gender memory, category guessing of age is governed by the conditional probability a(old | female) in Fig. 3. Note that these conditional guessing parameters are computed from the parameters for subgroup guessing according to Bayesian probability laws, so that they are not independent parameters.

A.2. Model structure

The basic structure of the model for crossed categories consists of five processing trees, one for each source of a statement. The sources are the four subgroups of speakers resulting from crossing the two binary category dimensions (i.e., young women, young men, old women, and old men) plus the class of new items (distracters). Accordingly, the data matrix for the response frequencies is enlarged relative to the standard model for simple categories (Klauer & Wegener, 1998), since there are now four instead of two categorical sources and corresponding assignments. Assignment decisions can thus be coded into 29 response cells (see Table in Appendix A.2). The cell on the lower left of Table in Appendix A.2 remains empty, as there cannot be a correct speaker for new items.

Fig. 3 shows the processing tree for statements made by a speaker from the subcategory of old women. It depicts the processing events leading to the six response categories specified in the first row of Table in Appendix A.2. Analogous processing trees correspond to the other rows of Table in Appendix A.2. The response categories are displayed as rectangular boxes on the right hand side of the graph. The model structure is analogous to that assumed in the model for simple categorization (Klauer & Wegener, 1998).

An item will be recognized as old with probability D, and its speaker will be remembered with probability c, so that the statement will be assigned to the correct person. In case the correct person is not remembered with probability $1 - c$, there might be subgroup memory

Data matrix of the model for crossed dichotomous categories

Source	Assignment to					
	The correct speaker	An incorrect speaker from the set of				The set of new items
		Old women	Young women	Old men	Young men	
Old women	1	2	3	4	5	6
Young women	7	8	9	10	11	12
Young men	13	14	15	16	17	18
Old men	19	20	21	22	23	24
New	n.a.	25	26	27	28	29

Note. n.a., not applicable. The response categories are numbered from 1 to 29.

with probability *d*(age, gender), exclusive age memory *d*(age, not gender), exclusive gender memory *d*(not age, gender), or no category memory at all *d*(not age, not gender). In the event of subgroup memory, the correct subcategory is retrieved and the correct speaker can still be guessed with a fixed probability of $1/n$ (here 1/3, since there are three discussants in each subgroup), resulting in a correct assignment, just as if the speaker had been remembered. If the correct speaker is not guessed, both category memberships are correctly remembered, but the speaker is not, a within-age–within-gender error. In the event of no category memory at all (*d*(not age, not gender)), stereotypical expectancies associated with the different subgroups may dispose participants to prefer one subgroup over others in assigning the statement to a speaker based on statement content. These guessing tendencies are quantified by the subgroup guessing parameters a. If the correct category, here that of old women, is guessed with probability *a*(old, female), the correct speaker can still be guessed with probability $1/n$, leading to a correct assignment, whereas a within-age–within-gender error occurs otherwise. If another subgroup is guessed, the assignment will be a between-categories error with respect to one or both dimensions.

If there is exclusive age memory with probability *d*(age, not gender), the perceiver remembers that the speaker was an old person, but cannot retrieve his or her gender. The above-mentioned stereotypical expectancies may again lead participants to prefer one of the gender categories among the old speakers in assigning the given statement. These preferences are quantified by the conditional guessing parameter *a*(female | old). If the correct gender category is guessed with probability *a*(female | old), the response can again be correct (with probability $1/n$) or a within-age–within-gender error (with probability $1 - 1/n$). Otherwise, a within-age–between-gender error occurs. The case of exclusive gender memory occurs with probability *d*(not age, gender), and leads to analogous assignment processes as the case of exclusive age memory. Finally, consider the case that the statement is not recognized as old ($1 - D$). It can then be guessed with probability b to have been presented in the discussion phase, reflecting response bias in item recognition. Since it then has to be assigned to someone, category guessing and person guessing processes take place in the same way as was just described for the case of complete absence of category memory. It is thus reasonably assumed that source recognition (category or person memory) cannot occur if the statement itself is not recognized in the first place. In case the item status is wrongly guessed to be new, the statement is falsely classified as a distracter (a false rejection) with probability $1 - b$.

Structurally equivalent processing trees account for the assignment of statements made by speakers from the other three subgroups. The processing tree for distracters is much simpler, as there are no memory processes involved in assigning items that were not presented. A distracter is correctly detected as being new with probability D_N. This kind of auto-noetic process (Strack & Bless, 1994) can be illustrated by inferences such as: "If this statement had been presented, I would certainly remember it, therefore it is new." If the item is not detected as new with probability $1 - D_N$, item status (old versus new) is guessed. With probability $(1 - b)$, the item can still be guessed to be new. In case it is erroneously guessed to be old (with probability b), the following processing events are characterized by the same subtree as follows the event "item guessed old" in Fig. 3: A speaker from a certain subcategory is chosen with preferences governed by the a parameters.

Note that in this model, processes of person memory are considered before processes involving category memory. The category memory parameters are thereby defined as conditional probabilities of retrieving category membership given that the speaker of a statement recognized as old cannot be remembered. Thus, category salience is assessed on the basis of only those recognized statements for which the speakers cannot be remembered. Assessing category memory conditional on the event that the speaker cannot be remembered is an inherent feature of the "Who said what?" paradigm that focuses on confusions between speakers. Note that the strength of category memory as assessed in the paradigm is nevertheless independent of memory for the individual speaker. If memory for individuating information becomes better, the basis for assessing category memory becomes smaller (in terms of the proportion of recognized statements without speaker memory), but what the strength of category memory may be is not in any way restrained. In fact, Klauer and Wegener (1998) separately manipulated factors affecting person memory and category salience and found independent and orthogonal effects on the category memory and the person memory parameters. When the discrimination of speakers was made more difficult, the person memory parameters decreased, but there were no significant effects on the category memory parameters (Exp. 5). Conversely, when category salience was manipulated, there were no significant effects on the measure of person memory, but the category memory parameters were strongly affected (Exp. 3). Assessing category memory conditional on the event that the speaker cannot be remembered does imply, however, that we cannot separately assess category memory conditional on the event that the speaker can be identified, and if one is not willing to assume that the two conditional probabilities (of category memory given speaker memory versus given no speaker memory) are equal, statements on this second component of category memory cannot be made.

Related multinomial models of source-monitoring for stimuli with multidimensional source information have been developed and successfully employed in the area of illusory correlations (Meiser & Hewstone, 2001) as well as with abstract, non-social stimuli (Meiser & Bröder, 2002). Advantages of this method over traditional procedures of analyzing data in the "Who said what?" paradigm are, first, that it avoids serious problems arising from item memory being less than perfect, as elaborated in Appendix B. Second, it overcomes interpretability problems by de-confounding, in an empirically validated manner, the relative contributions of the different processes that underlie the observed response frequencies (Klauer & Wegener, 1998) and thereby provides more direct measures of the psychological processes addressed in theories and substantive hypotheses. In particular, reconstructive category guessing can be assessed separately from category memory. Third, subgroup-level differences in the patterns of social categorization can be detected and examined which is not possible in the traditional analysis. Fourth and perhaps most importantly, the model makes explicit the assumptions upon which it is based and subjects them to an empirical examination through the model's goodness-of-fit test, whereas the assumptions underlying classical measures and their interpretations usually remain implicit and untested.

A.3. Restrictions and tests of the model used in Experiments 1–4

The entire model as described above already contains a number of methodological and substantial assumptions. First of all, the following technical restrictions had to be implemented for reasons of mathematical identifiability. To avoid overparametrization, that is, to guarantee that the model parameters can be estimated, it is necessary to anchor one D_N—parameter for distracter detection to at least one of the *D*-parameters for old items and to set the *b*-parameters for guessing the item status equal across all conditions (see Bayen, Murnane, & Erdfelder, 1996; Klauer & Wegener, 1998). In addition, it has to be assumed that the probabilities to guess category membership are the same regardless of whether an item was remembered or not remembered and guessed to be old; similarly, that the conditional probabilities of guessing category membership along one of the category dimensions are the same whether there is (exclusive) memory for the other dimension or not. These different processes contribute to the same response cells and cannot be distinguished in the data. Neither are there a-priori theoretical reasons to assume differences among these processes, nor empirical data suggesting such differences, and we therefore consider these necessary assumptions to be defensible. The ensemble of these assumptions is tested by the model's goodness-of-fit test.

All parameters except for the fixed probability of person guessing and, as just discussed, guessing of the item status, were allowed to vary over experiments. In addition, we implemented a number of substantial assumptions in order to keep the model as simple and as focused on the hypotheses under investigation as possible. The only parameters that were allowed to differ with respect to kind of statement were those assessing category guessing because stereotype-based reconstructive guessing is based on statement content. We tested for effects of speakers' category membership and kind of statement on the memory parameters—as would emerge, e.g., due to (in-)consistency effects—by letting them vary as a function of these factors. There were no such effects in any aspect of memory, as none of the accordingly relaxed models yielded an improvement of fit (for item recognition $\chi^2(28) = 27.88, p = .47$; for person memory $\chi^2(28) = 18.15, p = .90$; and for category-memory $\chi^2(84) = 87.19, p = .38$). Therefore, all memory parameters could be restricted to be equal over these two factors. Note that the absence of effects of speaker category also implies that there were no substantial ingroup–outgroup effects with respect to participants' age in these memory parameters.

To test for ingroup–outgroup effects of participants' gender, the model was fitted with separate parameters for the data from male and from female participants and with memory parameters that were allowed to vary as a function of the speakers' category memberships. Then we tested, separately for the different kinds of memory parameters and the category guessing parameters, whether setting equal the parameters over participants' gender would lead to a significant loss in model fit. In these many significance tests, there were only two significant results. Item memory parameters differed between male and female participants in the gender-visual study and the gender-verbal study. Inspection of the parameter estimates revealed that female participants showed somewhat better item memory uniformly over statements from all subgroups in these studies. Ingroup–outgroup effects would, however, manifest as cross-over interactions between participants' and discussants' gender, and there was no such pattern in the estimates. Ingroup–outgroup-effects are in fact rarely found in the "Who said what?"-paradigm (see Hewstone et al., 1991). As discussed in the General discussion, we attribute this to the fact that participants' evaluations play little role in the paradigm.

A.4. Restrictions and tests of the model used in Experiment 5

For Experiment 5, some of the restrictions just described did not hold (see also results section of Exp. 5). It seemed plausible that the more extreme photographs led to differential memory effects according to speakers' age category membership. Allowing person memory as well as category memory to differ according to speakers' age yielded a satisfactory fit.

Departing from this model version, we tested if model fit in Experiment 5 could be further improved by allowing for additional effects of congruency or incongruency, that is, for an interaction of speakers' category membership with kind of statement. This was neither the case for item recognition ($\chi^2(7) = 5.66, p = .42$) nor for person memory ($\chi^2(6) = 9.66, p = .14$) or category memory ($\chi^2(18) = 21.85, p = .24$), and the according restrictions (equality over kind of statement and speakers' gender category) were maintained. The final model used in the analysis of Experiment 5 differs from the one used in Experiments 1–4 only in that person memory and category memory are also allowed to vary as a function of the speakers' age category membership. As for Experiments 1–4, we also checked for ingroup–outgroup effects with respect to participants' gender, but found no evidence for such effects, although there was a non-significant tendency for female participants' item memory to be better overall than male participants' item memory.

Appendix B

In the traditional analysis of data from "Who said what?" tasks with crossed categories, confusions are classified in a two by two table according to whether they are within or between categories with respect to each dimension of categorization. The four cells of this table can accordingly be termed WW (e.g., within gender and within ethnicity), WB (e.g, within gender and between ethnicity), BW, and BB confusions, respectively, and thus confusion frequencies are ordered by a two-factorial design with error type (within versus between) along the first dimension and error type along the second dimension as factors. An interaction of the two factors such that WW confusions are proportionally over-represented (after a routine correction of WW confusions which we assume to have taken place throughout this discussion; cf. Stangor et al., 1992) is taken as evidence for dependencies in memory. Stangor et al. (1992) have lucidly argued that this analysis is inadequate:

> To appreciate this, consider an extreme case in which 90% of the overall sex errors are within-sex and 90% of the overall race errors are within-race. The expected errors in each of the cells of the resulting 2×2 matrix can be predicted by multiplying the row and column proportion marginals. In this case, one would expect that 81% of the overall errors would be within-race–within-sex, that 1% of errors would be between-race–between sex and that 18% (9% + 9%) of the errors would be within one category and between the other category.[] Such a pattern of means would produce a statistical interaction, even if categorization were only occurring on the basis of the two categories independently. (p. 211)

They go on to propose to estimate the expected number of WW confusions, under independence, as the product of the proportion of WB confusions among the WB and BB confusions times BW confusions among the BW and BB confusions times the total number of errors (i.e., $WB/(WB + BB) * BW/(BW + BB) * (WW + WB + BW + BB)$). Evidence for subgroup formation in memory is given if the estimated number of WW confusions falls short of the observed number. Clearly, this analysis more closely approximates a test for independence of the two dimensions of categorization.

Nevertheless, the analysis remains problematic. Using the modified version of the "Who said what?" paradigm with distracters (Klauer & Wegener, 1998), we typically find that a substantial percentage of discussion statements, 25% or more, is not remembered as having been presented in the discussion. For such statements, participants have little choice other than to assign the statement randomly to a speaker (setting aside the problem of systematic guessing bias as in reconstructive category guessing for the moment), whereas there will often be partial category memory for one or the other category distinction given statement memory. Thus, when there is no statement memory, the response is non-diagnostic with respect to assessing category memory and is equally likely to be classified as a WW, WB, BW, or BB confusion, whereas responses based on statement memory can also reflect partial category memory and are more likely to be classified as WW than as BW or WB confusions and least likely to become BB confusions to the extent to which there is in fact category memory. As a consequence, when item memory is less than perfect, the different kinds of confusions differ with respect to the proportion of non-diagnostic responses that stem from blind guessing in the absence of item memory and those that reflect partial memory.

Due to this confounding with item memory, estimating the proportion of WW confusions from ratios involving only the WB, BW, and BB confusions as proposed by Stangor et al. (1992) underestimates the extent of partial memory, because many responses classified as WB, BW, and BB will not permit an assessment of partial memory: Item memory was not given in the first place. This downward bias in estimating WW confusions leads to an overestimate of the extent of subtyping. For example, suppose that of 100 statements, an average of 24 statements are not remembered as having been part of the discussion and are randomly assigned leading one to expect an average of 6 confusions of each type (after correction of the number of WW confusions). Of the remaining 76 statements, let an average of 12 responses be correct identifications of the speaker, leaving 64 statements with item memory and speaker confusion. Assume that in 75% of these cases, there is memory for age and independently in 75% of these cases memory for gender. Multiplying row and column marginals leads one to expect an average of $36 = 64 * .75 * .75$ of these to be WW confusions,

$12 = 64 * .75 * .25$ WB confusions, 12 BW confusions, and $4 = 64 * .25 * .25$ BB confusions. Adding the 6 confusions of each type caused by blind guesses in the absence of statement memory, there will be an average of 42 WW, 18 WB, 18 BW, and 10 BB confusions. However, according to the above formula, the estimated number of WW confusions amounts to only $88 * (18/(18+10) * 18/(18+10)) = 36.4$, suggesting erroneously that there is evidence for subtyping in such data.

Another instance of this problem is given by the traditional tests for exclusive memory for a given dimension of categorization. Exclusive memory, or independent memory as it is often called, is said to be given when there are more within than between confusions for one dimension among those confusions that were between-errors on the other dimension, that is, WB and BB, or BW and BB are contrasted. Again, the proportion of non-diagnostic responses for which there was no item memory in the first place is likely to be greater in these contrasts than in the parallel contrasts on the basis of within-errors on the other dimension, that is in contrasting WW and BW, or WW and WB, respectively. This leads to an underestimate of exclusive memory for the category dimension under scrutiny. The model analyses of the present paper partial out effects of item memory and category guessing and thereby correct for this and additional problems.

References

Arcuri, L. (1982). Three patterns of social categorization in attribution memory. *European Journal of Social Psychology, 12*, 271–282.

Bayen, U. J., Murnane, K., & Erdfelder, E. (1996). Source discrimination, item detection, and multinomial models of source-monitoring. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22*, 197–215.

Batchelder, W. H., & Riefer, D. M. (1999). Theoretical and empirical review of multinomial processing tree modeling. *Psychonomic Bulletin and Review, 6*, 57–86.

Biernat, M., & Vescio, T. K. (1993). Categorization and stereotyping: Effects of group context on memory and social judgment. *Journal of Experimental Social Psychology, 29*, 166–202.

Bodenhausen, G. V., & Macrae, C. N. (1998). Stereotype activation and inhibition. In R. S. Wyer (Ed.), *Stereotype activation and inhibition. Advances in social cognition, Vol. 11* (pp. 1–52). Mahwah, Erlbaum: Erlbaum.

Brewer, M. B. (1979). In-group bias in the minimal group situation. A cognitive-motivational analysis. *Psychological Bulletin, 86*, 307–324.

Brewer, M. B. (1988). A dual process model of impression formation. In T. K. Srull, & R. S. Wyer (Eds.), *Advances in social cognition* (Vol. 10, pp. 1–36). Hillsdale, NJ: Erlbaum.

Brewer, M. B., Dull, V., & Lui, L. (1981). Perceptions of the elderly: Stereotypes as prototypes. *Journal of Personality and Social Psychology, 41*, 656–670.

Brewer, M. B., Ho, H.-K., Lee, J.-Y., & Miller, N. (1987). Social identity and social distance among Hong Kong school children. *Personality and Social Psychology Bulletin, 13*, 156–165.

Brewer, M. B., & Lui, L. N. (1989). The primacy of age and sex in the structure of person categories. *Social Cognition, 3*, 262–274.

Crisp, R. J., Ensari, N., Hewstone, M., & Miller, N. (in press). A dual-route model of crossed categorization effects. Manuscript in preparation. To appear in W. Stroebe & M. Hewstone (Eds.), *European Review of Social Psychology (Vol. 13)*. Hove, UK: Psychology Press.

Crisp, R. J., & Hewstone, M. (1999). Subcategorization of physical stimuli: Category differentiation and decategorization processes. *European Journal of Social Psychology, 29*, 665–672.

Crisp, R. J., & Hewstone, M. (2000). Multiple categorization and social identity. In D. Capozza, & R. Brown (Eds.), *Social identity theory: Trends in theory and research* (pp. 149–166). Beverly Hills: Sage.

Crisp, R. J., Hewstone, M., & Rubin, M. (2001). Does multiple categorization reduce intergroup bias? *Personality and Social Psychology Bulletin, 27*, 76–89.

Deschamps, J.-C. (1977). Effect of crossing category memberships on quantitative judgement. *European Journal of Social Psychology, 7*, 517–521.

Deschamps, J.-C., & Doise, W. (1978). Crossed category membership in intergroup relations. In H. Tajfel (Ed.), *Differentiation between social groups* (pp. 141–158). Cambridge: Cambridge University Press.

Doise, W. (1978). *Groups and individuals: Explanations in social psychology*. New York: Cambridge University Press.

Gawronski, B., Ehrenberg, K., Banse, R., Zukova, J., & Klauer, K. C. (2003). It's in the mind of the beholder: The impact of stereotypic associations on category-based and individuating impression formation. *Journal of Experimental Social Psychology, 39*, 16–30.

Gilbert, D. T., & Hixon, J. G. (1991). The trouble of thinking: Activation and application of stereotypic beliefs. *Journal of Personality and Social Psychology, 60*, 509–517.

Hewstone, M., Hantzi, A., & Johnston, L. (1991). Social categorization and person memory: The pervasiveness of race as an organizing principle. *European Journal of Social Psychology, 21*, 517–528.

Hewstone, M., Islam, M. R., & Judd, C. M. (1993). Models of crossed categorization and intergroup relations. *Journal of Personality and Social Psychology, 64*, 779–793.

Hewstone, M., Macrae, C. N., Griffiths, R., Milne, A. B., & Brown, R. (1994). Cognitive models of stereotype change: (5) Measurement, development and consequences of subtyping. *Journal of Experimental Social Psychology, 30*, 505–526.

Klauer, K. C., & Meiser, T. (2000). A source-monitoring analysis of illusory correlations. *Personality and Social Psychology Bulletin, 26*, 1074–1093.

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

Klauer, K. C., Wegener, I., & Ehrenberg, K. (2002). Perceiving minority members as individuals: The effects of relative group size in social categorization. *European Journal of Social Psychology, 32*, 223–245.

Lenton, A. P., Blair, I. V., & Hastie, R. (2001). Illusions of gender: Stereotypes evoke false memories. *Journal of Experimental Social Psychology, 37*, 3–14.

Lepore, L., & Brown, R. (1997). Category and stereotype activation: Is prejudice inevitable. *Journal of Personality and Social Psychology, 72*, 275–287.

Lippmann, W. (1922). *Public opinion*. New York: Harcourt Brace.

Macrae, C. N., Bodenhausen, G. V., & Milne, A. B. (1995). The dissection of selection in person perception: Inhibitory processes in social stereotyping. *Journal of Personality and Social Psychology, 69*, 397–407.

Meiser, T., & Bröder, A. (2002). Memory for multidimensional source information. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28*, 116–137.

Meiser, T., & Hewstone, M. (2001). Crossed categorization effects on the formation of illusory correlations. *European Journal of Social Psychology, 31*, 443–466.

Migdal, M. J., Hewstone, M., & Mullen, B. (1998). The effects of crosses categorization on intergroup evaluations: A Meta-analysis. *British Journal of Social Psychology, 37*, 303–324.

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

Park, B., Wolsko, C., & Judd, C. M. (2001). Measurement of subtyping in stereotype change. *Journal of Experimental Social Psychology, 37*, 325–332.

Pendry, L. F., & Macrae, C. N. (1996). What the disinterested perceiver overlooks: Goal-directed social categorization. *Personality and Social Psychology Bulletin, 22*, 249–256.

Sinclair, L., & Kunda, Z. (1999). Reactions to a black professional: Motivated inhibition and activation of conflicting stereotypes. *Journal of Personality and Social Psychology, 77*, 885–904.

Stangor, C., & Lange, J. E. (1993). Mental representation of social groups: Advances in understanding stereotypes and stereotyping. In M. P. Zanna (Ed.), *Advances in Experimental Social Psychology* (Vol. 26, pp. 357–416). New York: Academic Press.

Stangor, C., Lynch, L., Duan, C., & Glass, B. (1992). Categorization of individuals on the basis of multiple social features. *Journal of Personality and Social Psychology, 62*, 207–218.

Stangor, C., & McMillan, D. (1992). Memory of expectancy-congruent and expectancy-incongruent information: A review of the social and social developmental literature. *Psychological Bulletin, 111*, 42–61.

Strack, F., & Bless, H. (1994). Memory for nonoccurrences: Metacognition and presuppositional strategies. *Journal of Memory and Language, 33*, 203–217.

Tajfel, H. E. (1978). *Differentiation between social groups: Experiments in the social psychology of intergroup relations*. London: Academic Press.

Tajfel, H., Flament, C., Billig, M., & Bundy, R. F. (1971). Social categorization and intergroup behaviour. *European Journal of Social Psychology, 11*, 149–178.

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*, 778–793.

Urban, L. M., & Miller, N. (1998). A theoretical analysis of crossed categorization effects: A meta-analysis. *Journal of Personality and Social Psychology, 74*, 894–908.

van Knippenberg, A., van Twuyver, M., & Pepels, J. (1994). Factors affecting social categorization processes in memory. *British Journal of Social Psychology, 33*, 419–431.

van Knippenberg, A., & Dijksterhuis, A. (2000). Social categorization and stereotyping: A functional perspective. In W. Stroebe, & M. Hewstone (Eds.), *European review of social psychology* (Vol. 11, pp. 105–144). Chichester: Wiley.

van Rijswijk, W., & Ellemers, N. (2002). Context effects on the application of stereotype content to multiple categorizable targets. *Personality and Social Psychology Bulletin, 28*, 90–101.

van Twuyver, M., & van Knippenberg, A. (1995). Social categorization as a function of priming. *European Journal of Social Psychology, 25*, 695–701.

van Twuyver, M., & van Knippenberg, A. (1998). Effects of group membership and identification on categorization and subtyping in memory. *European Journal of Social Psychology, 28*, 531–553.

Vanbeselaere, N. (1991). The different effects of simple and crossed categorizations: A result of the category differentiation process or of differential category salience?. In W. Stroebe, & M. Hewstone (Eds.), *European review of social psychology* (Vol. 2, pp. 247–278). Chichester: Wiley.

Vescio, T. K., Judd, C. M., & Kwan, V. S. L. (in press). The crossed categorization hypothesis: Evidence of reductions in the strength of categorization but not intergroup bias. *Journal of Experimental Social Psychology*. Manuscript submitted for publication.

Wegener, I. (2000). *Soziale Kategorien im situativen Kontext: Kognitive Flexibilität bei der Personenwahrnehmung.* [Social categories in the situative context: Cognitive flexibility in person perception.] Unpublished doctoral dissertation. University of Bonn, Germany.

Wittenbrink, B., Judd, C. M., & Park, B. (2001a). Evaluative versus conceptual judgments in automatic stereotyping and prejudice. *Journal of Experimental Social Psychology, 37*, 244–252.

Wittenbrink, B., Judd, C. M., & Park, B. (2001b). Spontaneous prejudice in context variability in automatically activated attitudes. *Journal of Personality and Social Psychology, 81*, 815–827.

Wyer, R. S., & Srull, T. K. (1989). Human cognition in its social context. *Psychological Review, 93*, 322–359.

Zárate, M. A., & Smith, E. R. (1990). Person categorization and stereotyping. *Social Cognition, 8*, 161–185.