

Kotzur, Patrick F. et al.

Article — Published Version

‘Society thinks they are cold and/or incompetent, but I do not’:
Stereotype content ratings depend on instructions and the social
group's location in the stereotype content space

British Journal of Social Psychology

Provided in Cooperation with:
WZB Berlin Social Science Center

Suggested Citation: Kotzur, Patrick F. et al. (2020) : ‘Society thinks they are cold and/or incompetent, but I do not’: Stereotype content ratings depend on instructions and the social group's location in the stereotype content space, British Journal of Social Psychology, ISSN 2044-8309, Wiley, Hoboken, Vol. 59, Iss. 4, pp. 1018-1042, <https://doi.org/10.1111/bjso.12375>

This Version is available at:
<https://hdl.handle.net/10419/228484>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



<https://creativecommons.org/licenses/by-nc/4.0/>



‘Society thinks they are cold and/or incompetent, but I do not’: Stereotype content ratings depend on instructions and the social group’s location in the stereotype content space

Patrick F. Kotzur^{1*} , Susanne Veit^{2,3}, Annika Namyslo^{1†},
Mirka-Alicia Holthausen^{4‡}, Ulrich Wagner⁴ and Ruta Yemane²

¹University of Osnabrück, Germany

²Berlin Social Science Center, Germany

³German Center for Integration and Migration Research, Berlin, Germany

⁴Philipps-University of Marburg, Germany

Stereotype content researchers have grown accustomed to ask participants how ‘society’ views social groups to tap into culturally shared stereotype content and to reduce social desirability bias (J Person Soc Psychol, 82, 2002, 878). However, methodological and theoretical considerations raise questions about this common practice, and stereotype content researchers have also asked for participants’ personal perspective on social groups in the past. Nonetheless, how and whether stereotype content model scores empirically differ as a function of the instructed perspective remains questionable and to date untested. Thus, we investigated whether and, if so, how stereotype content results are affected when instructing participants to evaluate social groups from society’s versus their personal perspective. Across three experiments (Study 1: $N = 301$; Study 2: $N = 126$; Study 3: $N = 1,221$), latent mean comparisons indicated that results regarding stereotype content ratings are affected by the instructed perspective (society’s vs. personal) contingent on the social group’s location in the stereotype content space: Stereotype content ratings were more negative when participants were asked to provide society’s perspective on social groups compared to their own perspective, but only on an already depreciated stereotype content dimension. The number of possible comparisons across experimental conditions was substantially reduced, since preconditions for these analyses were not met. Given our methodological and theoretical considerations and empirical corroborations that the instructed perspective does affect results, we encourage a discussion on how to best measure culturally shared stereotype content and propose aggregating stereotype content scores from participants’ personal perspective to the cultural level.

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

*Correspondence should be addressed to Patrick F. Kotzur, Department of Psychology, University of Osnabrück, Seminarstraße 20, 49074 Osnabrück, Germany (email: pakotzur@uos.de).

†Present address: Psychology Department, University of Koblenz-Landau Germany

‡Present address: Psychology Department, University Duisburg-Essen Germany

The stereotype content model (SCM; Fiske, Cuddy, Glick, & Xu, 2002) is a popular model of the social perception of social groups. Researchers have applied the model to map stereotype content of social groups within societies (e.g., Asbrock, 2010; Burkley, Durante, Fiske, Burkley, & Andrade, 2017; Bye *et al.*, 2014; Clausell & Fiske, 2005; Eckes, 2002; Janssens, Verkuyten, & Khan, 2015; Kotzur, Friehs, Asbrock, & Van Zalk, 2019; Sadler *et al.*, 2012) and to investigate regional within-country and cross-country differences in the endorsement of stereotype content (e.g., Binggeli, Krings, & Sczesny, 2014; Cuddy *et al.*, 2009; Durante *et al.*, 2013, 2017; Stanciu, Cohrs, Hanke, & Gavreliuk, 2017). In brief, stereotype content of social groups is based on two basic dimensions of social perception: warmth, the social groups' 'potential harm or benefit of the target group's goals' (Cuddy, Fiske, & Glick, 2007, p. 632) to the ingroup's goals, and competence, the 'degree to which the group can effectively enact those goals' (Cuddy *et al.*, 2007, p. 632). A group's warmth levels are determined by threat and competition perceptions, whereas groups' competence levels are predicted by their status (Binggeli *et al.*, 2014; Fiske *et al.*, 2002; Kervyn, Fiske, & Yzerbyt, 2015). Warmth and competence perception jointly give rise to contemptuous (low warmth/low competence), envious (low warmth/high competence), and paternalistic (high warmth/low competence) outgroup evaluations, as well as admiration towards allied and ingroups (high warmth/high competence; Cuddy *et al.*, 2009). Stereotype content has further implications for how perceivers intend to behave towards groups; warm or competent groups elicit facilitating intentions, cold, or incompetent harming intentions (Cuddy *et al.*, 2007).

Despite – or because of – its continuous popularity *as is*, researchers have identified various methodological and conceptual avenues to advance the model. Among other things, concurrent discussions focus on whether or not stereotype content scales' preconditions for statistical analyses are met (e.g., Friehs, Kotzur, Zöller, Wagner, & Asbrock, 2020; Kotzur, Friehs, *et al.*, 2019), and whether additional or alternative (sub-) dimensions may be needed to adequately capture stereotype content (e.g., Abele, Hauke, Peters, Louvet, Szymkow, & Duan, 2016; Koch, Imhoff, Dotsch, Unkelbach, & Alves, 2016; Leach, Ellemers, & Barreto, 2007; Stanciu *et al.*, 2017). We contribute a further facet to these theoretical and methodological considerations: The implications of whether participants are asked to provide stereotype content ratings from their own or society's perspective.

In the SCM literature, it has become customary to ask individuals to indicate how 'society' views social groups (e.g., Asbrock, 2010; Bye *et al.*, 2014; Fiske *et al.*, 2002; Kotzur, Friehs, *et al.*, 2019), usually using the following question: 'As viewed by society, how [item] are members of this group?' (Fiske *et al.*, 2002, p. 891). Researchers in this area put forward mostly two reasons. First, they aim to tap into *culturally shared* stereotype content, that is, into how social groups are perceived in a given societal context (Fiske, 2017; Fiske *et al.*, 2002). Indeed, according to the theory, stereotype content is meant to capture cultural stereotypes that are universally shared among individuals within a given context (see Fiske *et al.*, 2002, p. 881). Asking individuals how 'society' views social groups is meant to achieve this goal (Fiske *et al.*, 2002). However, researchers from other fields in social psychology operationalize culture-related constructs differently. In cultural value research, for instance, individual responses aimed to measure personal endorsement of values are often aggregated to represent cultural-level scores (Schwartz, 2011). Using this logic, individuals' *personal* values are used to find the 'true' average *cultural* values within a context. Conceptually, then, aggregating individual scores to one mean score of a higher entity (e.g., culture) represents the shared cultural construct (although individual deviations are still expected), which could be used to map social groups within

and across societies. Similar approaches have been used in other fields of study in psychology and social sciences (e.g., Ariely & Davidov, 2011; Gelfand *et al.*, 2011; Hofstede, 1980). According to this line of reasoning, stereotype content literature using this instruction assesses individuals' *perceived* culturally shared stereotype content, which may not necessarily correspond with personal or 'truly' culturally shared stereotype content.

Second, many researchers argue that instructing how 'society' views social groups diminishes social desirability bias (e.g., Fiske *et al.*, 2002), that is, the tendency to deviate from one's 'true' answer to answer items in more socially acceptable ways (Lavrakas, 2008). However, this argument contradicts the rationale for the procedure outlined above. This is because this argument draws on the assumption that participants provide their *personal* and 'true' views with less inhibition when being asked for 'society's' perspective. As such, this argument suggests that instructions are not chosen to tap into *culturally shared* perceptions of groups, as the first argument implies, but instead into participants' 'true' personal perceptions. Moreover, the argument that SCM scores are overall less biased using this technique becomes less convincing when one considers that this technique may fall prey to other kinds of biases. To refer to but one example, such assessments may be governed by the false consensus bias (Ross, Greene, & House, 1977), that is, the tendency to project one's own beliefs onto others, thereby overestimating how much other people share one's own beliefs.

The question of how the instructed perspective affects SCM results gains further urgency when considering the many ways researchers have assessed warmth and competence. For instance, some authors explicitly asked for participants' personal assessment of groups when providing SCM scores (e.g., Kotzur *et al.*, 2017). Such deviations in instructions date back even earlier, including the original study (Fiske *et al.*, 2002, Study 3), when the researchers asked 'How [item] are members of this group?', not explicitly asking participants to take 'society's' perspective.

Another potential issue with the instructed perspective is raised by the authors of the original study, based in the United States of America:

Regarding meaning of responses, were participants reporting the culture's, their group's, or their own personal stereotypes and prejudices? The questionnaire at the outset emphasized the project's interest in American society and at the top of each page instructed participants to answer "as viewed by society." However, as the questionnaire went on, participants may have forgotten these instructions and begun to respond as individuals or group members [...]. In retrospect, we might have phrased those items differently (Fiske *et al.*, 2002, p. 898).

Thus, how to formulate the instructions in SCM research to measure the construct at hand has been identified as a challenge from early on.

In sum, we attest contradicting goals between the two main rationales to ask for society's perspective (i.e., measuring cultural vs. less biased personal stereotype content), different measuring practices that do not ask explicitly for society's or individuals' perspective, and the suspicion that even when instructing participants to take society's perspective, participants may fail to do so. However, as it becomes apparent from the reviewed literature, how and whether the instructed perspective actually does affect SCM scores remains questionable and to date untested. With three experimental studies, two utilizing a between-subject design (Studies 1 and 3) one a within-subject design (Study 2), we aim at contributing insights to see whether the instructed perspective does affect SCM scores. Significant differences would indicate that the instructed perspective does make a

difference, suggesting a debate, including a critical reflection of current practices in the current stereotype content literature, is warranted. Moreover, irrespective of its outcomes, the results of this study additionally contribute to our understanding of how past and future SCM studies using differing instructions relate to one another and join others that aim to contribute to a discussion on measurement issues of the SCM (e.g., Friehs *et al.*, 2020; Kotzur, Friehs, *et al.*, 2019).

Study I

The first study was designed to investigate how warmth and competence scores are affected by instructing participants to either providing society's versus their personal perspective on six social groups using a between-subject design.

Methods

Sample, procedure, and measures

We followed Kotzur, Friehs *et al.*'s (2019) recommendations to analyse SCM data in a latent variable framework, since it allows for accounting for measurement error (Kline, 2010) and for formally testing whether preconditions for mean comparisons can be established (measurement invariance; Vandenberg & Lance, 2000). Boomsma (1982, 1985) recommended a minimum sample size of 100 to 200 for latent variable modelling. We recruited 301 participants (94.7% university students, 72.8% female, 27.2% male, $M_{age} = 23.44$, $SD_{age} = 6.29$) in Germany via university-wide email lists and social media postings to participate in a brief online study. Participants were randomly allocated to one of two conditions (society's perspective condition, $n = 152$; personal perspective condition, $n = 149$). Participants in the society's perspective condition received the instructions used in most of Fiske *et al.*'s (2002) original studies: 'Responding to the questions we are not interested in your personal opinion, but rather how you think the majority of society would rate various social groups'. In the personal view condition, the instruction read: 'Responding to the questions we are interested in your personal opinion how you would rate various social groups'.¹

Thereafter, we asked participants on a scale from 1 = *not at all* to 5 = *completely* how warm, friendly, and well-intentioned for warmth, and competent, independent, and competitive for competence² (Asbrock, 2010; Eckes, 2002; Kotzur *et al.*, 2017; Kotzur, Friehs, *et al.*, 2019; Kotzur, Schäfer, & Wagner, 2019) either society (society's perspective condition) or themselves (personal perspective condition) rated six social groups from different quadrants in the SCM space (Asbrock, 2010): athletes and students (high warmth/high competence), elderly (high warmth/low competence), Muslims (mid-warmth/mid-competence), the rich (low warmth/high competence), and the homeless (low warmth/low competence).³ After providing demographical data, participants were thanked and had the opportunity to get compensated for their participation with partial course credit.

¹ In German: 'Bei der Beantwortung der Fragen geht es nicht um Ihre persönliche Meinung, sondern darum, dass Sie angeben, wie die Mehrheit der Gesellschaft verschiedene soziale Gruppen Ihrer Perspektive nach einschätzen würde'. and 'Bei der Beantwortung der Fragen geht es um Ihre persönliche Meinung, wie Sie verschiedene gesellschaftliche Gruppen einschätzen'.

² In German: warm, sympathisch, gutmütig, kompetent, eigenständig, and konkurrenzfähig, respectively.

³ In German: Sportler, Studenten, alte Menschen, Muslime, Reiche, and Obdachlose, respectively.

Analysis strategy

Following the procedure of Kotzur, Friehs *et al.* (2019), we first assessed the general measurement model for each of the six social groups in each condition separately using confirmatory factor analyses. The latent warmth and competence factors were allowed to correlate, whereas no cross-loadings of items between factors were allowed. We used Schermelleh-Engel, Moosbrugger, and Müller's (2003) criteria for adequate model fit: $\chi^2/df \leq 3$; root mean standard error of approximation (RMSEA) $\leq .08$; standardized root mean square residual (SRMR) $\leq .10$; and comparative fit index (CFI) $\geq .95$.

Mean comparisons between experimental conditions require scalar measurement invariance. Using a step-up approach, we established scalar invariance across conditions for each group separately that had achieved acceptable model fit in both experimental conditions in the previous analysis step (Vandenberg & Lance, 2000; see also Figure 1). Finally, given scalar measurement invariance across experimental conditions, we inspected whether latent means of the warmth and competence factors differed significantly ($p_{two-tailed} \leq .05$) between conditions (society's perspective vs. personal perspective) to test whether these instructions lead to systematic differences in the levels of SCM ratings.

Results

All analyses were conducted in Mplus version 8, using robust maximum likelihood estimator (MLR) to account for multivariate non-normality, non-independence of observations, and missing data (Muthén & Muthén, 1998–2017).

Preliminary analyses

Following the procedure described above, we assessed the general measurement model for each of the six social groups in each condition separately using confirmatory factor analyses. Only the models for Muslims reached adequate model fit in both experimental conditions separately to be considered for further analyses (see Table 1). Scalar invariance

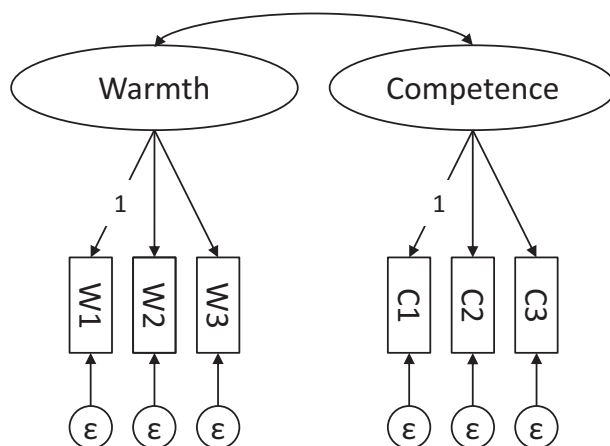


Figure 1. Structural model specified in Study 1 to conduct latent mean comparisons across experimental conditions (society's perspective vs. personal perspective). W1-3 = warmth indicators, C1-3 = competence indicators. ϵ = error term.

Table 1. Study 1: Model fit indices of measurement models for each social group and condition separately

	Personal perspective condition						Society's perspective condition						Comparable?				
	<i>n</i>	χ^2_{cor}	<i>df</i>	<i>p</i>	χ^2_{cor}/df	SRMR	<i>n</i>	χ^2_{cor}	<i>df</i>	<i>p</i>	χ^2_{cor}/df	RMSEA		CFI	SRMR		
Athletes	119	20.58	8	.008	2.57	.12	.94	.05	122	12.30	8	.138	1.53	.07	.96	.05	No
Elderly	118	22.07	8	.005	2.76	.12	.91	.07	120	27.65	8	.001	3.46	.14	.84	.10	No
Homeless	121	26.11	8	.001	3.26	.14	.90	.06	119	29.77	8	<.001	3.72	.15	.90	.05	No
Muslims	119	14.37	8	.073	1.80	.08	.97	.04	122	13.64	8	.092	1.71	.08	.98	.04	Yes
Rich	119	18.75	8	.016	2.34	.11	.95	.06	122	16.00	8	.042	2.00	.09	.94	.05	No
Students	122	15.88	8	.044	1.99	.09	.96	.06	121	11.44	8	.178	1.43	.06	.96	.06	No

Note. We categorized those social groups as comparable across experimental conditions that produced measurement models with satisfactory model fit in both experimental conditions separately. Satisfactory model fit was indicated if all of the following requirements were fulfilled: $\chi^2_{cor}/df \leq 3$; RMSEA $\leq .10$; and CFI $\geq .95$ (Schermelleh-Engel, et al., 2003).

Table 2. Study 1: Results of measurement invariance testing across experimental conditions for social groups that previously produced measurement models with satisfactory model fit in both experimental conditions separately

	<i>n</i>	χ^2_{cor}	<i>df</i>	<i>p</i>	χ^2_{cor}/df	RMSEA	CFI	SRMR	$\Delta\chi^2_{\text{cor}}$	Δdf	<i>p</i>
Muslims											
Configural	241	28.18	16	.030	1.76	.08	.97	.04			
Metric	241	29.74	20	.074	1.49	.06	.98	.04	0.22	4	.994
Scalar	241	31.63	24	.137	1.32	.05	.98	.04	1.39	4	.846

Note. $\Delta\chi^2_{\text{cor}}$, Δdf , and *p* refer to chi-square difference test results with Satorra–Bentler scaling correction (more freely estimated model vs. more constrained model).

Table 3. Latent mean comparison of warmth and competence between society's perspective and personal perspective condition

	<i>n</i>	Warmth			Competence		
		Estimate	SE	<i>p</i>	Estimate	SE	<i>p</i>
Muslims	241	0.36	0.10	<.001	0.28	0.09	.003

Note. Society's perspective condition served as a reference group (latent mean = 0).

in the model containing the warmth and competence factors for Muslims across experimental conditions could be established (see Table 2), allowing for meaningful latent mean comparisons.

Main analyses

In order to assess how asking for society's versus one's personal perspective affects SCM results, we inspected whether Muslims' latent mean values of warmth and competence in the society's perspective condition (fixed to 0, reference condition) were significantly different from the latent mean values of warmth and competence in the personal perspective condition. Both warmth ($\kappa = 0.36$, $p < .001$) and competence ($\kappa = 0.28$, $p = .003$) were significantly higher in the personal perspective condition, indicating that asking for individuals' personal perspective on Muslims leads participants to evaluate them more positively on both dimensions than when asking for society's perspective (see Table 3).

Discussion

Study 1 showed that Muslims received systematically more positive evaluations on both warmth and competence dimensions when asked for one's personal assessment. These findings indicate that, first, people do differentiate between personal and societal perspective, which serves as a first empirical indication that instructed perspectives matter. Second, our findings suggest that studies using instructions consistent with most of Fiske *et al.*'s (2002) reported studies asking for society's perspective may have systematically lower warmth and competence scores than studies that instructed participants to report their own perspective (e.g., Kotzur *et al.*, 2017).

Unfortunately, we were only able to compare one out of six groups across experimental conditions, due to poor model fit of the remaining measurement models, indicating that the specified structural model consistent with the theoretical model did not reflect the empirical reality. This is despite our usage of items that have been previously established in the German context and used in many SCM studies in Germany (e.g., Eckes, 2002; Asbrock, 2010; Kotzur *et al.*, 2017; Kotzur, Friehs, *et al.*, 2019; Kotzur, Schäfer, *et al.*, 2019; see also Friehs *et al.*, 2020). Although loss of data applying more rigorous statistical analyses is to be expected (Kotzur, Friehs, *et al.*, 2019), it remains to be seen whether the results that we found for Muslims can be generalized to other social groups as well. We conducted Study 2 to address this question.

Study 2

The second study was designed to replicate the findings of Study 1 regarding the question how warmth and competence scores are affected by instructing participants to either providing society's versus their personal views with an improved study design. These improvements concerned (a) increasing the number of included social groups to nine to increase the likelihood that more measurement models with adequate model fit will emerge, (b) increasing the number of items per construct to increase reliability of scales and allowing for ad hoc adjustments (e.g., excluding items that underperform or testing for partial invariance; Kotzur, Friehs, *et al.*, 2019), and (c) using a within-subject design to reduce strain on our constrained participant pool, while matching power requirements (minimal $N = 100$ to 200 Boomsma, 1982, 1985).

Methods

Sample, procedure, and measures

We recruited 126 (Wave 1) participants (81.7% university students, 71.4% female, 27.8% male, 0.8% other, $M_{age} = 27.51$, $SD_{age} = 11.87$) in Germany via university-wide email lists and online social network postings for a brief online study. Data collection was completed in two waves 1 week apart. All participants were asked to provide both society's (in one wave) and their personal assessment (in another wave) of nine social groups (within-subject factor: instructed perspective [1 = personal, 2 = society]). In each wave, participants were asked to provide one of these perspectives for all groups. The order in which participants were asked to provide society's view or their personal view in the first or second wave was balanced to avoid order effects (54% were presented with items asking for participants' personal perspective in the first wave). Twenty-five participants could not be matched between waves. Univariate chi-square and *t*-tests revealed no significant differences on demographic variables between those that dropped out and those that continued the study beyond Wave 1, indicating that dropout occurred non-systematically ($ps > .05$). Missing values were again estimated using MLR.

We used the identical instructions and items and scales to measure warmth and competence as in Study 1. To increase the reliability of the scales, we translated and included two additional items per scale from Fiske *et al.*'s (2002) original study: sincere and tolerant for warmth, and confident and intelligent for competence.⁴ We asked participants to rate nine social groups from different quadrants in the SCM space

⁴ In German: *aufrichtig, tolerant, zuversichtlich, and intelligent, respectively.*

(Asbrock, 2010): physicians, Germans (high warmth/high competence), elderly, housewives (high warmth/low competence), Turks (mid-warmth/mid-competence), the rich, feminists (low warmth/high competence), and the homeless and jobless (low warmth/low competence).⁵ After providing demographical data, participants were thanked and invited to take part in the second wave. At the end of the questionnaire at the second wave, participants had the opportunity to get compensated for their participation with partial course credit.

Analysis strategy

We tested our hypothesis in a latent variable framework again, using the same model fit criteria and procedure as in Study 1, with two exceptions. First, since we now had five indicators per construct available, we ran a series of measurement models for warmth and competence factors separately per condition and group (identical to those in Study 1, except with more indicators per latent factor) to identify and exclude items that underperformed. We applied an iterative exclusion procedure, in a way that we excluded one item at a time, selecting the indicator that performed poorest with the highest number of social groups guided by modification indices (potential χ^2 -improvements ≥ 10), until the highest possible number of models achieved satisfactory model fit. Only those models with groups that achieved acceptable model fit in both experimental conditions were included in the next analysis steps.

Moreover, participants' instructed perspective was a within-subject factor in this study, and we were thus interested in mean comparisons between conditions *within* participants. Therefore, we next specified one latent warmth and competence factor per group from a personal perspective (Warmth_p and Competence_p) and one warmth and competence factor from society's perspective (Warmth_s and Competence_s). Latent factors were allowed to correlate. Moreover, error terms of items with identical wordings were assumed to share variance and thus were allowed to correlate across conditions (see also Figure 2). After establishing scalar invariance, we inspected whether latent means of the warmth and competence factors differed significantly ($p_{two-tailed} \leq .05$) between conditions (personal perspective vs. society's perspective) to test whether the instructions lead to systematic differences in the SCM ratings.

Results

Preliminary analyses

Following the procedure described above, we assessed the general measurement model for each of the nine social groups in each condition separately. Using all five indicators per construct, none of the models reached adequate fit in both conditions. Fourteen of 18 models produced modification indices (see Tables S1 and S5). After removing 'warm' from the warmth factor (8 times indicated as problematic in the previous models, Tables S2 and S5), 'confident' from the competence factor (8 times indicated as problematic in the previous models, Tables S3 and S5), and 'competitive' from the competence factor (6

⁵In German: Ärzte, Deutsche, alte Menschen, Hausfrauen, Türken, Reiche, Feministinnen, Obdachlose, and Arbeitslose, respectively.

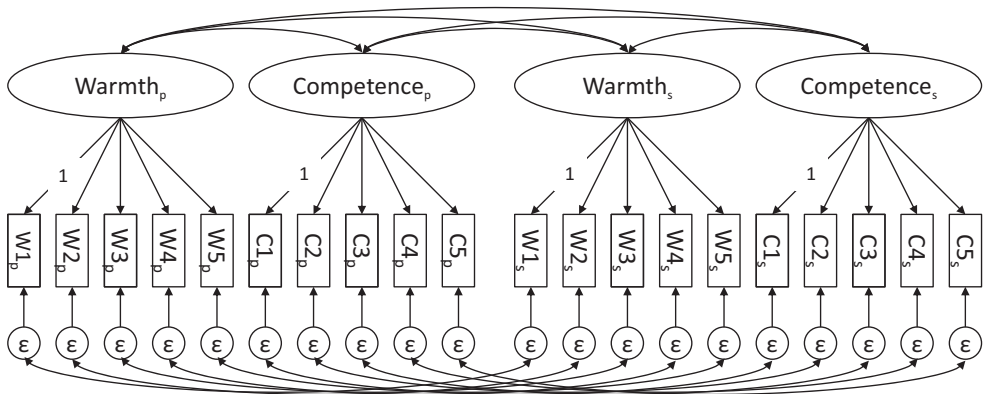


Figure 2. Structural model specified in Study 2 to conduct latent mean comparisons across experimental conditions. Number of indicators varies depending on the social group compared. W1-5_p = warmth indicators of the personal perspective condition, C1-5_p = competence indicators of the personal perspective condition, W1-5_s = warmth indicators of the society's perspective condition, C1-5_s = competence indicators of the society's perspective condition. ε = error term.

times indicated as problematic in the previous models, Tables S4 and S5), only two measurement models indicated modification indices (Germans and jobless from society's perspective; see Table 4). Excluding these indicators, the models for elderly, rich, and Turks reached adequate fit in both experimental conditions to be considered for further analyses. Since the measurement models for jobless and physicians had satisfactory fit in both conditions with more items per construct, which deteriorated with further exclusions of items, we continued to consider these groups with more items per construct in the next analysis step.⁶ Partial or full scalar invariance in the model containing the warmth and competence factors across experimental conditions could be established for elderly, jobless, rich, and Turks (see Table 5), allowing for meaningful latent mean comparisons.

Main analyses

In order to assess how asking for personal versus society's perspective on social groups affects SCM results, we inspected whether the latent mean values of warmth and competence in the society's perspective condition (fixed to 0, reference condition) were significantly different from the latent mean values of warmth and competence in the personal perspective condition (see Table 6). Compared to when asked for society's perspective, participants indicated that they personally attributed both more warmth and competence to jobless, $\kappa = 0.57, p = .002$ and $\kappa = 0.80, p < .001$, as well as Turks, $\kappa = 0.44, p < .001$ and $\kappa = 0.36, p < .001$, respectively. Moreover, participants reported to personally attribute more competence to elderly, $\kappa = 0.14, p = .043$, and more warmth to the rich, $\kappa = 0.30, p < .001$.

⁶ For jobless, we considered a model in which warm, confident, and competitive were excluded (see Table S5). For physicians, we considered a model in which warm and confident were excluded (see Tables S3).

Table 4. Study 2: Model fit indices of measurement models for each social group and condition separately, excluding the items warm, confident, competitive, and tolerant

	Personal perspective condition						Society's perspective condition										
	<i>n</i>	χ^2_{cor}	<i>df</i>	<i>p</i>	χ^2_{cor}/df	RMSEA	CFI	SRMR	<i>n</i>	χ^2_{cor}	<i>df</i>	<i>p</i>	χ^2_{cor}/df	RMSEA	CFI	SRMR	Comparable?
Elderly	116	4.94	8	.764	0.62	.00	1.00	.04	111	12.24	8	.141	1.53	.07	.96	.05	Yes
Feminists	116	11.23	8	.189	1.40	.06	.99	.04	111	23.90	8	.002	2.99	.13	.91	.06	No
Germans	116	11.28	8	.186	1.41	.06	.97	.04	111	16.92	8	.031	2.12	.10	.92	.07	No
Homeless	116	15.03	8	.059	1.88	.09	.92	.06	111	6.11	8	.635	0.76	.00	1.00	.03	No
Housewives	116	13.10	8	.109	1.64	.07	.96	.05	111	16.90	8	.031	2.11	.10	.92	.06	No
Jobless	116	13.16	8	.107	1.64	.08	.98	.04	111	19.20	8	.014	2.40	.11	.94	.05	No
Physicians	116	9.66	8	.289	1.21	.04	.99	.03	111	23.36	8	.003	2.92	.13	.92	.07	No
Rich	116	12.45	8	.132	1.56	.07	.98	.05	111	11.55	8	.172	1.44	.06	.98	.06	Yes
Turks	116	9.94	8	.269	1.24	.05	.99	.04	111	9.01	8	.341	1.13	.03	.99	.04	Yes

Note. We categorized those social groups as comparable across experimental conditions that produced measurement models with satisfactory model fit in both experimental conditions separately. Satisfactory model fit was indicated if all of the following requirements were fulfilled: $\chi^2_{cor}/df \leq 3$; RMSEA $\leq .08$; SRMR $\leq .10$; and CFI $\geq .95$ (Schermelleh-Engel, et al., 2003).

Table 5. Study 2: Results of measurement invariance testing across experimental conditions for social groups that previously produced measurement models with satisfactory model fit in both experimental conditions separately

	<i>n</i>	χ^2_{cor}	<i>df</i>	<i>p</i>	χ^2_{cor}/df	RMSEA	CFI	SRMR	$\Delta\chi^2_{cor}$	Δdf	<i>p</i>
Elderly											
Configural	126	53.66	42	.107	1.28	.05	.96	.07			
Metric	126	54.34	46	.187	1.18	.04	.97	.08	1.47	4	.832
Scalar	126	57.32	50	.222	1.15	.03	.97	.08	3.11	4	.539
Jobless											
Configural	126	84.20	64	.046	1.32	.05	.96	.08			
Metric	126	98.02	69	.012	1.42	.06	.94	.09	13.26	5	.021
Partial metric	126	84.80	68	.082	1.25	.04	.97	.08	0.94	4	.919
Scalar	126	91.64	71	.050	1.29	.05	.96	.08	7.26	3	.064
Physicians											
Configural	126	145.79	90	.000	1.62	.07	.91	.08			
Rich											
Configural	126	54.81	42	.089	1.30	.05	.97	.06			
Metric	126	59.08	46	.093	1.64	.05	.97	.06	4.40	4	.355
Scalar	126	66.51	50	.059	1.33	.05	.96	.06	7.64	4	.106
Turks											
Configural ^a	126	48.22	42	.236	1.15	.03	.98	.06			
Metric ^a	126	60.56	46	.074	1.32	.05	.96	.09	12.68	4	.013
Partial metric ^a	126	53.14	45	.189	1.18	.04	.98	.07	4.95	3	.176
Scalar ^a	126	59.84	48	.118	1.25	.04	.97	.09	6.92	3	.074

Note. For elderly, rich, and Turks, we considered a model in which the items warm, confident, competitive, and tolerant were excluded (see Table 4). For jobless, we considered a model in which the items warm, confident, and competitive were excluded (see Table S5). For physicians, we considered a model in which the items warm and confident were excluded (see Table S3).

^aIncludes Heywood case. We refrained from testing for metric and scalar invariance for physicians, since the previous model did not reach configural invariance. $\Delta\chi^2_{cor}$, Δdf , and *p* refer to chi-square difference test results with Satorra–Bentler scaling correction (more freely estimated model vs. more constrained model).

Table 6. Study 2: Latent mean comparison of warmth and competence between society’s perspective and personal perspective condition

	Warmth			Competence		
	Estimate	SE	<i>p</i>	Estimate	SE	<i>p</i>
Elderly	−0.02	0.06	.770	0.14	0.07	.043
Jobless	0.57	0.18	.002	0.80	0.15	<.001
Rich	0.30	0.08	.000	−0.10	0.07	.134
Turks	0.44	0.07	<.001	0.36	0.06	<.001

Note. Society’s perspective condition served as a reference group (latent mean = 0).

Discussion

In Study 2, four out of nine comparisons across conditions were possible (compared to one out of six in Study 1). The pattern of results of these possible comparisons

jointly painted a more nuanced picture than the results of Study 1. Going beyond the implications of Study 1, these findings indicate that a more pronounced depreciation of social groups occurs when asking for society's perspective compared to when one is asked for one's personal perspective; however, which of the stereotype content dimensions was more depreciated varied as a function of the group's location in the two-dimensional SCM space. Rich, depreciated on warmth (Asbrock, 2010; Fiske *et al.*, 2002), were even more depreciated on warmth in the society's perspective condition than in the personal perspective condition (yet not on competence). Elderly, depreciated on competence (Asbrock, 2010; Fiske *et al.*, 2002), were even more depreciated on competence in the society's view condition (yet not on warmth). Jobless and Turks, both depreciated on both warmth and competence (Asbrock, 2010), were even more depreciated when participants were instructed to report society's compared to their personal perspective. Thus, participants tended to provide less harsh judgements on the stereotype content dimension that is commonly depreciated when asked about their personal perspective. These results and interpretations are consistent with the results of Study 1, where Muslims, depreciated on both warmth and competence (Asbrock, 2010), were even more depreciated on both stereotype content dimensions in the society's perspective condition compared to the personal perspective condition. When comparing SCM findings using these differing instructions then, it seems likely that SCM scores using the society's perspective instruction produced more dispersed results, since scores on the depreciated dimension may tend to be more extremely depreciated than when participants are instructed to provide their own judgement (see Figure 3). Again, and

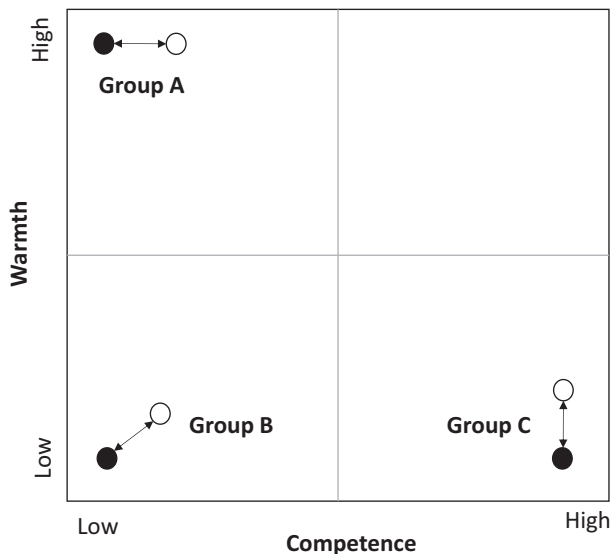


Figure 3. Graphical representation of shifts in mean levels as a function of perspective (personal vs. society's) and location of the social group in the stereotype content space (example group A: high warmth/low competence; example group B: low warmth/low competence; or example group C: low warmth/high competence). Black dots represent social groups evaluated by society's perspective and white dots by one's personal perspective.

even more importantly, people do differentiate between personal and societal perspective, which serves as an indicator that instructed perspectives matter.

Like many SCM studies before us, both Study 1 and Study 2 have been based on student samples. Students tend to be more liberal and more educated than the general adult population, both factors potentially shaping their evaluation of social groups (Hodson & Busseri, 2012; Webster, Burns, Pickering, & Saucier, 2014) which may limit the generalizability of our findings. Moreover, the low performance of the scales we used in Studies 1 and 2 to measure stereotype content limited the amount of meaningful comparisons, and thus conclusions we could draw, based on these studies. We conducted Study 3 to address these points.

Study 3

The purpose of the third study was to extend the findings of Studies 1 and 2 by investigating our research questions in a heterogeneous adult sample using a different scale to measure stereotype content. To this end, we used data from an online survey that was originally collected for a different purpose (Veit, & Yemane, 2020), namely the evaluations of Germans and various migrant groups in Germany from different countries of origin.⁷

Sample, procedure, and measures

The initial sample consisted of 2,323 participants from a commercial online panel in Germany (Clickworker) who participated in a survey in exchange for a small monetary reward (2.00 € for less than 15 min of their time). After providing demographical data and evaluating some photographs irrelevant to the present study, participants evaluated a subset of up to three migrant groups (of 38 different origins in total: That is, they either evaluated Germans or migrants in Germany who themselves or whose parents migrated to Germany from one out of 37 foreign countries). Importantly, participants were randomly allocated to one of two conditions. They rated groups either from their personal or from society's perspective, resulting in a between-subject design like Study 1.

Studies 1 and 2 jointly suggested that the social groups' location in the SCM space matters as to which stereotype content dimension will be more depreciated in the societal perspective condition. Thus, in order to revisit these findings, we included only those participants in our analyses that evaluated social groups whose location in the SCM space had already been established in the German context (see below). Moreover, adopting Lee and Fiske's (2006) procedure when studying migrant groups in the SCM framework, we only included majority-member participants in our analyses (i.e., participants with

⁷ The survey was conducted with the aim of merging its results with the data collected in two large-scale field experiments on ethnic discrimination in hiring (see Lancee et al., 2019; Veit, & Yemane, 2018). It consisted of two parts: first, a short evaluation of the application photographs that were used in the two field experiments (photo evaluation, not relevant for the present study but mentioned for the sake of completeness), and second, a survey on respondents' stereotypical views of the 38 social groups of different countries of origin that were considered in the field experiments (stereotype survey). The 38 groups were clustered into three larger groups (i.e., sets of social groups). To ensure that each participant evaluated at least one group he or she was familiar with, we randomly assigned from each set one group to participants, respectively. However, because of a mistake in code for the assignment of application photographs in the photo survey, we had to stop the survey after some weeks of data collection and slightly changed the survey design. To compensate for missing observations in the photo survey, we merged all three sets of social groups in the stereotype survey and randomly assigned one group out of the 38 social groups to each participant. As a result of this change in the design, 1,368 participants were asked to evaluate three groups and 955 participants were asked to evaluate only one group.

German nationality). These inclusion criteria resulted in our final sample of 1,221 participants (level of education: 31% master's (or equivalent) or higher, 39% bachelor's (or equivalent) or vocational training certificate, 18% higher education entrance qualification, 12% General Certificate of Secondary Education or lower, 0% other, and 1% missing; 43% female, 45% male, 0% other, and 12% missing; $M_{\text{age}} = 40.74$, $SD_{\text{age}} = 10.68$), with $n = 583$ in the personal perspective and $n = 638$ in the society's perspective condition.

Similar to Studies 1 and 2, participants in the society's perspective were instructed the following: 'We would like to ask you to evaluate to what extent people in Germany associate particular characteristics with particular societal groups in Germany'. In the personal perspective condition, the instruction read: 'We would like to ask you to evaluate to what extent you personally associate particular characteristics with particular societal groups in Germany'.⁸ Attempting to measure warmth and competence more effectively, participants were asked to evaluate on 7-point semantic differentials consisting of opposing adjective pairs based on Asbrock (2010) and Fiske *et al.* (2002) how *dislikable–likable*, *untrustworthy–trustworthy*, *cold–warm*, and *threatening–benevolent* for warmth, and *incompetent–competent*, *lazy–laborious*, *unreliable–reliable* and *uneducated–highly educated* for competence⁹ either society (society's perspective condition) or participants themselves (personal perspective condition) rated one to three social groups, of which we included the following 13 in our analyses: migrants of Albanian, Bulgarian, Greek, Italian, Moroccan, Pakistani, Polish, Romanian, Russian, Turkish, and Egyptian origin as low to mid-warmth/low to mid-competence migrant groups (Fröhlich & Schulte, 2019), and migrants of Chinese origin and Germans as low warmth/high competence groups (Fröhlich & Schulte, 2019).¹⁰ Subsample sizes per social group and condition varied between 45 and 81 and were thus lower than is usually recommended (Boomsma, 1982, 1985).

Analysis strategy

We used the same analytical strategy as in our other study utilizing a between-subject design (Study 1), while also adopting the model fit optimization procedure for underperforming measurement models we have used in Study 2. Moreover, since subsample size per condition and social group was lower than recommended (Boomsma, 1982, 1985), we additionally investigated post hoc whether power was sufficient to detect meaningful differences between experimental conditions. To replicate the pattern of previous findings, low to mid-warmth/low to mid-competence groups (migrants of Albanian, Bulgarian, Egyptian, Greek, Italian, Moroccan, Pakistani, Polish, Romanian, Russian, and Turkish origin; Fröhlich & Schulte, 2019) should receive lower warmth and competence scores in the society's compared to personal perspective condition. Low warmth/high competence groups (migrants of Chinese origin and Germans; Fröhlich & Schulte, 2019) should receive lower warmth scores in the society's perspective condition compared to personal perspective condition.

⁸ In German: 'Wir wollen Sie nun bitten einzuschätzen, inwiefern Menschen in Deutschland bestimmte Eigenschaften mit bestimmten Bevölkerungsgruppen in Deutschland verbinden' and 'Wir wollen Sie nun bitten einzuschätzen, inwiefern Sie persönlich bestimmte Eigenschaften mit bestimmten Bevölkerungsgruppen in Deutschland verbinden', respectively.

⁹ In German: *Unsympathisch–sympathisch*, *nicht vertrauenswürdig–vertrauenswürdig*, *kaltherzig–warmherzig*, and *bedrohlich–wohlwollend* for warmth, and *inkompetent–kompetent*, and *faul–fleißig*, *unzuverlässig–zuverlässig*, and *ungebildet–hochgebildet* for competence.

¹⁰ In German: *Migranten aus Albanien, Bulgarien, Griechenland, Italien, Marokko, Pakistan, Polen, Rumänien, Russland, der Türkei, Ägypten, China, and Deutsche*, respectively.

Results

Preliminary analyses

First, we assessed the general measurement model for each of the thirteen social groups in each condition separately. Using all four indicators per construct, four groups reached adequate model fit in both experimental conditions (Germans, as well as Albanian, Polish, and Turkish migrants). Four of 26 models produced modification indices (see Tables S6 and S8). After removing 'cold-warm' from the warmth factor (4 times indicated as problematic in the previous models, Tables S7 and S8)¹¹ and 'uneducated-highly educated' from the competence factor (1 time problematic in the previous models),¹² no measurement models indicated modification indices (see Tables 7 and S8). Excluding these indicators, Albanian, Italian, Polish, and Turkish migrants reached adequate fit in both experimental conditions to be considered for further analyses. Since the measurement models for Germans, Bulgarian, Moroccan, and Pakistani migrants had satisfactory fit in both conditions with the 'uneducated-highly educated' item included, yet not without it, we continued to consider these groups with this item in the next analysis step.

We were able to establish partial or full scalar invariance across experimental conditions for the models of Albanian, Bulgarian, Moroccan, Pakistani, Polish, and Turkish migrants (see Table 8), allowing for meaningful latent mean comparisons.

Main analyses

In order to assess how asking for personal versus society's perspective on social groups affects SCM results, we inspected again whether the latent mean values of warmth and competence in the society's perspective condition were significantly different from the latent mean values of warmth and competence in the personal perspective condition (see Table 9). Compared to when asked for society's perspective, participants indicated that they personally attributed both more warmth and competence to Albanian migrants, $\kappa = 0.55$, $p = .020$ and $\kappa = 0.52$, $p = .004$, Bulgarian migrants, $\kappa = 0.75$, $p = .004$ and $\kappa = 0.59$, $p = .001$, and Turkish migrants, $\kappa = 0.58$, $p = .020$ and $\kappa = 0.45$, $p = .036$. These results corroborated the findings from Studies 1 and 2 that groups depreciated on both warmth and competence are rated less benevolently on both stereotype content dimensions in the society's perspective condition. Pakistani migrants received higher warmth scores in the personal compared to the societal view condition, $\kappa = 0.49$, $p = .004$. The remaining differences remained non-significant, yet pointed in the expected direction.

Next, we performed post hoc power analyses to assert whether power issues may have prevented us from detecting significant instruction effects in these instances. Following Muthén and Muthén (2002), we used Monte Carlo simulations with 10,000 replications and MLR estimator based on estimates provided by the models above,

¹¹ 'cold-warm' and 'untrustworthy-trustworthy' emerged both four times as problematic in the previous model. Thus, we continued to fit models also for groups that we did not include in our main analyses, and found that 'cold-warm' emerged fourteen times as problematic, whereas untrustworthy-trustworthy did eleven times. Thus, we removed the overall more problematic item 'cold-warm' from the models.

¹² 'Unreliable-reliable', 'lazy-laborious', 'uneducated-highly educated', and 'incompetent-competent' emerged once as problematic in the previous model. We followed the same strategy as above, finding that across all models, 'uneducated-highly educated' was the most problematic item (ten times), followed by 'unreliable-reliable' (seven times), 'incompetent-competent' (six times), and 'lazy-laborious' (five times). Thus, we removed 'uneducated-highly educated' from the models.

Table 7. Study 3: Model fit indices of measurement models for each group and condition separately, excluding the items cold-warm and uneducated-highly educated

	Personal perspective condition						Society's perspective condition										
	n	χ^2_{cor}	df	p	χ^2_{cor}/df	RMSEA	CFI	SRMR	n	χ^2_{cor}	df	P	χ^2_{cor}/df	RMSEA	CFI	SRMR	Comparable?
Albanian	58	1.67	8	.990	0.21	.00	1.00	.02	73	3.27	8	.916	0.41	.00	1.00	.02	Yes
Bulgarian	66	7.69	8	.464	0.96	.00	1.00	.03	53	15.32	8	.053	1.92	.13	.95	.04	No
Chinese	63	18.18	8	.020	2.27	.14	.91	.05	81	16.34	8	.038	2.04	.11	.96	.05	No
Egyptian	66	13.90	8	.084	1.74	.11	.97	.03	62	11.14	8	.194	1.39	.08	.99	.03	No
German	60	12.29	8	.139	1.54	.09	.95	.05	57	10.03	8	.263	1.25	.07	.97	.05	No
Greek	58	11.02	8	.201	1.38	.08	.99	.03	45	12.62	8	.126	1.58	.11	.95	.07	No
Italian	55	9.15	8	.330	1.14	.05	.99	.04	53	9.87	8	.275	1.23	.07	.98	.06	Yes
Moroccan	45	11.22	8	.190	1.40	.10	.99	.02	49	4.10	8	.848	0.51	.00	1.00	.02	No
Pakistani	47	2.79	8	.947	0.35	.00	1.00	.02	72	12.16	8	.144	1.52	.09	.98	.03	No
Polish	53	8.33	8	.402	1.04	.03	1.00	.04	59	6.11	8	.635	0.76	.00	1.00	.03	Yes
Romanian	57	26.19	8	.001	3.27	.20	.86	.06	61	3.86	8	.870	0.48	.00	1.00	.02	No
Russian	57	3.81	8	.874	0.48	.00	1.00	.03	74	12.37	8	.135	1.55	.09	.98	.04	No
Turkish	50	5.73	8	.680	0.72	.00	1.00	.02	73	11.18	8	.192	1.40	.07	.98	.04	Yes

Note. We categorized those groups as comparable across experimental conditions that produced measurement models with satisfactory model fit in both experimental conditions separately. Satisfactory model fit was indicated if all of the following requirements were fulfilled: $\chi^2_{cor}/df \leq 3$; RMSEA $\leq .08$; SRMR $\leq .10$; and CFI $\geq .95$ (Schemmelleh-Engel, et al., 2003).

Table 8. Study 3: Results of measurement invariance testing across experimental conditions for groups that previously produced measurement models with satisfactory model fit in both experimental conditions separately

	<i>n</i>	χ^2_{cor}	<i>df</i>	<i>p</i>	χ^2_{cor}/df	RMSEA	CFI	SRMR	$\Delta\chi^2_{\text{cor}}$	Δdf	<i>p</i>
Albanian											
Configural	131	4.66	16	.997	0.29	.00	1.00	.02			
Metric	131	11.51	20	.932	0.58	.00	1.00	.08	9.41	4	.052
Scalar	131	14.19	24	.942	0.59	.00	1.00	.07	2.82	4	.589
Bulgarian											
Configural	119	28.44	26	.337	1.09	.04	.99	.04			
Metric	119	40.52	31	.118	1.31	.07	.97	.11	13.56	5	.019
Partial metric	119	34.67	30	.255	1.16	.05	.99	.08	6.91	4	.141
Scalar	119	38.54	34	.272	1.13	.05	.99	.09	3.81	4	.432
German											
Configural	117	32.07	26	.191	1.23	.06	.97	.05			
Metric	117	45.99	31	.041	1.48	.09	.92	.14	12.37	5	.030
Partial metric	117	40.48	30	.096	1.35	.08	.94	.12	7.74	4	.101
Italian											
Configural	108	18.88	16	.275	1.18	.06	.99	.05			
Metric	108	24.36	20	.227	1.22	.06	.98	.10	5.49	4	.241
Scalar	108	27.37	24	.288	1.14	.05	.99	.11	2.78	4	.595
Moroccan											
Configural	94	23.90	26	.582	0.92	.00	1.00	.02			
Metric	94	26.23	31	.710	0.85	.00	1.00	.04	2.19	5	.823
Scalar	94	33.57	36	.585	0.93	.00	1.00	.05	7.66	5	.176
Pakistani											
Configural	120	35.71	26	.097	1.37	.08	.98	.03			
Metric	120	41.08	31	.106	1.33	.07	.98	.06	5.18	5	.823
Scalar	120	50.07	36	.060	1.39	.08	.97	.07	9.58	5	.088
Polish											
Configural	112	14.38	16	.571	0.90	.00	1.00	.03			
Metric	112	18.81	20	.534	0.94	.00	1.00	.07	4.83	4	.305
Scalar	112	23.82	24	.472	0.99	.00	1.00	.09	5.09	4	.278
Turkish											
Configural	123	17.17	16	.375	1.07	.03	1.00	.03			
Metric	123	21.86	20	.348	1.09	.04	1.00	.06	4.68	4	.321
Scalar	123	22.27	24	.563	0.93	.00	1.00	.07	0.55	4	.968

Note. For Germans and Bulgarian, Moroccan, and Pakistani migrants, we considered a model in which the item cold–warm was excluded (see Table S7). For Albanian, Italian, Polish, and Turkish migrants, we considered a model in which the items cold–warm and uneducated–highly educated were excluded (see Table 7). We refrained from testing for scalar invariance for Germans, since the partial metric model did not meet model fit requirements. We refrained from submitting Italian migrants to further analyses, since the model fit of the scalar model was below cut-off criteria. $\Delta\chi^2_{\text{cor}}$, Δdf , and *p* refer to chi-square difference test results with Satorra–Bentler scaling correction (more freely estimated model vs. more constrained model).

assuming normally distributed data with no missings. Overall, post hoc power estimates ranged from .14 to .99, with all power estimates below the .80 criterion (Cohen, 1992) for estimates that did not significantly differ between conditions, yet did point in the expected direction (.14–.73).

Table 9. Study 3: Latent mean comparison of warmth and competence between society's perspective and personal perspective condition

	Warmth				Competence			
	Estimate	SE	<i>p</i>	Power	Estimate	SE	<i>p</i>	Power
Albanian	0.55	0.24	.020	.96	0.52	0.18	.004	.99
Bulgarian	0.75	0.26	.004	.98	0.59	0.18	.001	.98
Moroccan	0.42	0.26	.112	.73	0.32	0.26	.223	.40
Pakistani	0.49	0.25	.004	.90	0.36	0.23	.118	.69
Polish	0.37	0.25	.145	.55	0.15	0.24	.526	.14
Turkish	0.58	0.25	.020	.97	0.45	0.21	.036	.94

Note. Society's perspective condition served as a reference group (latent mean = 0). Power refers to the post hoc power estimates obtained using Monte Carlo simulations with 10,000 replications and MLR estimator, assuming normally distributed and complete data (Muthén & Muthén, 2002).

Discussion

Study 3 provided further evidence that instructing participants to take their personal or society's perspective when providing SCM assessments leads to systematic differences in responses. It added to Studies 1 and 2 by drawing on a heterogeneous adult sample and by using a novel stereotype content scale consisting of semantic differentials based on items provided by previous literature (Asbrock, 2010; Fiske *et al.*, 2002). Six out of thirteen comparisons across conditions were possible, resulting in a similar ratio (46%) of obtainable comparisons compared to the other two studies of the present research (one out of six or 17% in Study 1, and four out of nine or 44% in Study 2). We interpret this result as a first step towards the development of a more reliable and valid German-language scale measuring stereotype content, yet have to acknowledge that more work needs to be done to ultimately achieve this goal. The pattern of results largely corroborated the ones from Studies 1 and 2. Again, a more pronounced depreciation of social groups emerged when asking for society's perspective compared to one's personal perspective, especially for the particularly depreciated stereotype content dimension for Albanian, Bulgarian, and Turkish migrants (depreciated more harshly on both warmth and competence; Asbrock, 2010; Fröhlich & Schulte, 2019). Pakistani migrants were more depreciated on warmth in the societal, compared to the personal perspective condition. Moreover, the remaining expected mean differences for Pakistani, Polish, and Moroccan migrants were in the expected direction, yet not significant. Post hoc power analyses suggested that a lack of adequate power may have been the culprit, given that the sample sizes were below conventional recommendations (e.g., Boomsma, 1982, 1985). In appreciation of these limiting factors, we suggest that these findings serve as initial evidence that, overall, the processes found in Studies 1 and 2 may also generalize to heterogeneous adult populations using different scales. We can thus conclude that instructed perspectives matter, also in more diverse samples and different measurement instruments.

General discussion

With the aims to tap into stereotype content that is culturally shared (Fiske, 2017; Fiske *et al.*, 2002) and to reduce social desirability concerns (Fiske *et al.*, 2002), SCM researchers have grown accustomed to ask how 'society' views social groups. Given a

contradiction between these two goals, potential other biases that may operate when asking for society's perspective (e.g., Ross *et al.*, 1977), existing research that asks for participants' personal view of social groups (e.g., Kotzur *et al.*, 2017), and early concerns about the implications of the instructed perspective (Fiske *et al.*, 2002), we conducted three experiments testing whether and how the instructed perspective affects stereotype content scores. Doing so, we also provided insights as to how past and future SCM studies using differing instructions relate to one another and join others' concurrent efforts that aim to add to a discussion on the measurement of SCM (e.g., Friehs *et al.*, 2020).

Our three experiments produced two key findings. Firstly, they showed consistently that the obtained results depended significantly on the instructed perspective with which social groups were asked to be evaluated (society's vs. one's own), in a way that groups tended to receive more negative evaluations when participants were asked to provide society's perspective. Secondly, Study 2 suggested that the stereotype content dimension that gets more negatively evaluated in this context depended on which dimension is depreciated more strongly and thereby on the social group's location in the two-dimensional SCM space. High competence/low warmth groups may be seen as less warm, low competence/high warmth groups as less competent, and low competence/low warmth groups as less competent and warm when asked for society's compared to one's personal perspective. As Studies 1 and 2 suggest, these findings apply to student samples. Study 3 provided initial evidence that these results may also extend to heterogeneous adult samples.

We have thus gathered empirical support for the idea that systematic differences emerge in SCM studies when using these differing instructions, in a way that SCM scores using the society's perspective instruction are likely to produce more dispersed and more negative results, since scores on the depreciated dimension may tend to be more extremely depreciated than when participants are instructed to provide their own judgement. These findings suggest that a debate regarding the used instructions in the stereotype content literature is warranted.

One suggestion that may address all theoretical and methodological issues that arise from asking for society's perspective on groups would be to follow other scholars in social psychology asking for participants' personal perspective and to aggregate these personal scores to the cultural level (Ariely & Davidov, 2011; Gelfand *et al.*, 2011; Hofstede, 1980; Schwartz, 2011). The resulting measures at the aggregate level represent cultural-level scores as envisioned by the original authors (Fiske, 2017; Fiske *et al.*, 2002). Second, applying this procedure would help to get rid of the – in our view – ambiguous role instructions play in eliciting participants' 'true', unbiased evaluation of groups. That is, asking for participants' personal perspective would leave less room for doubt as to whether participants provide their own assessment, or some – in potentially other ways biased – assessment on how others of a reference group view social groups. Therefore, we invite SCM researchers to consider the well-established procedures in other lines of social psychology to establish cultural-level scores (e.g., Ariely & Davidov, 2011; Gelfand *et al.*, 2011; Hofstede, 1980; Schwartz, 2011).

We see multiple ways future research could build on our findings. Firstly, we provided initial evidence of main effects of instructed perspective on warmth and competence ratings. Future research could extend this line of research by investigating further effects on additional components of the SCM and its extensions (emotions and action tendencies; Cuddy *et al.*, 2007).

Secondly, although we used established scales to measure warmth and competence in Studies 1 and 2 that have been used in a multitude of studies (e.g., Asbrock, 2010; Eckes, 2002; Kotzur *et al.*, 2017; Kotzur, Friehs, *et al.*, 2019; Kotzur, Schäfer, *et al.*, 2019), the number of possible comparisons was harshly reduced given the low performance of these scales (measurement models produced unsatisfactory fit in ten out of 15 cases across both of these studies, or 67%). Indeed, a reanalysis of 19 published SCM data sets in Germany mostly using these items revealed that measurement models with satisfactory model fit could not be established in 58 out of 94 cases (or 62%; Friehs *et al.*, 2020). Following Kotzur, Friehs *et al.*'s (2019) assessment, this reduction is attributable to the more rigorous statistical analyses in a latent variable modelling framework compared to previously published research (but see Janssens *et al.*, 2015; Kotzur, Friehs, *et al.*, 2019; Stanciu *et al.*, 2017), increasing the validity of our findings. Including more items and more groups (which both cannot be done indefinitely, of course), both strategies that we applied in Studies 2 and 3, may alleviate such problems in the short run. Future research could complement these strategies with smart study designs, such as planned missingness (e.g., Graham, Taylor, Olchowski, & Cumsille, 2006), to reduce the additional burden on individual participants. In the long run, we believe that it is imperative that researchers aim to improve the scale properties of warmth and competence, benefitting the entire SCM research community that relies on robust and well-performing measures. We provided a first attempt to that end in Study 3.

Thirdly, due to low scale performance with allied and ingroups (high warmth/high competence groups), we were unable to investigate how instructions affected their ratings. If our expectation based on the observed empirical pattern holds that only depreciated dimensions are more depreciated when asked for society's compared to one's personal perspective, no significant differences should emerge for such groups. Yet, only more research in this domain can provide more insights to that end.

Conclusion

Stereotype content model researchers often ask for 'society's' view on social groups. Given contradicting goals between the two main rationales to ask for society's perspective on groups measuring *cultural* versus less biased *personal* stereotype content, the existence of measuring practices that do not ask explicitly for society's or individuals' perspectives, and the suspicion that even when instructing participants to take society's perspective, participants may fail to do so, we tested whether the instructed perspective affects SCM scores. Over the course of three studies, we provided evidence that SCM ratings indeed differ between instructed perspectives (society's vs. personal) and that this difference depends on the social group's location in the stereotype content space. Specifically, social groups tended to be evaluated more negatively when participants were asked to provide society's perspective on social groups compared to their own perspective, but only on an already depreciated SCM dimension. We believe that our results, as well as theoretical and methodological arguments, warrant a debate regarding the ways we want to measure culturally shared stereotype content. As a viable way forward, we suggest that stereotype content researchers consider the well-established procedures in other lines of social psychology to establish cultural-level scores by asking for personal perception and aggregating them to the cultural level (e.g., Ariely & Davidov, 2011; Gelfand *et al.*, 2011; Hofstede, 1980; Schwartz, 2011).

Conflicts of interest

All authors declare no conflict of interest.

Authors' contribution

Patrick Ferdinand Kotzur: Conceptualization, Formal analysis, Investigation, Methodology, Project administration, Visualization, Writing – original draft, Writing – review & editing; Susanne Veit: Data curation, Writing – review & editing; Annika Namyslo: Data curation, Writing – review & editing; Mirka-Alicia Holthausen: Data curation, Writing – review & editing; Ulrich Wagner: Resources, Supervision, Writing – review & editing; Ruta Yemane: Data curation.

Data availability statement

Raw data of Studies 1, 2, and 3 are available upon request from the corresponding author; data from Studies 1 and 2 cannot be made publicly available, since participants were informed that data management would be controlled by the study authors at all times.

References

- Abele, A. E., Hauke, N., Peters, K., Louvet, E., Szymkow, A., & Duan, Y. (2016). Facets of the fundamental content dimensions: Agency with competence and assertiveness – Communion with warmth and morality. *Frontiers in Psychology*, *7*, 1810. <https://doi.org/10.3389/fpsyg.2016.01810>
- Ariely, G., & Davidov, E. (2011). Can we rate public support for democracy in a comparable way? Cross-national equivalence of democratic attitudes in the World Value Survey. *Social Indicators Research*, *104*, 271–286. <https://doi.org/10.1007/s11205-010-9693-5>
- Asbrock, F. (2010). Stereotypes of social groups in Germany in terms of warmth and competence. *Social Psychology*, *41*, 76–81. <https://doi.org/10.1027/1864-9335/a000011>
- Binggeli, S., Krings, F., & Sczesny, S. (2014). Perceived competition explains regional differences in the stereotype content of immigrant groups. *Social Psychology*, *45*, 62–70. <https://doi.org/10.1027/1864-9335/a000160>
- Boomsma, A. (1982). The robustness of LISREL against small sample sizes in factor analysis models. In K. G. Joreskog & H. Wold (Eds.), *Systems under indirect observation: Causality, structure, prediction (Part I)*. Amsterdam, Netherlands: North-Holland.
- Boomsma, A. (1985). Nonconvergence, improper solutions, and starting values in LISREL maximum likelihood estimation. *Psychometrika*, *50*, 229–242. <https://doi.org/10.1007/BF02294248>
- Burkley, E., Durante, F., Fiske, S. T., Burkley, M., & Andrade, A. (2017). Structure and content of Native American stereotypic subgroups: Not just (ig)noble. *Cultural Diversity and Ethnic Minority Psychology*, *23*, 209–219. <https://doi.org/10.1037/cdp0000100>
- Bye, H. H., Herrebrøden, H., Hjetland, G. J., Røyset, G. Ø., & Westby, L. L. (2014). Stereotypes of Norwegian social groups. *Scandinavian Journal of Psychology*, *55*, 469–476. <https://doi.org/10.1111/sjop.12141>
- Clausell, E., & Fiske, S. T. (2005). When do subgroup parts add up to the stereotypic whole? Mixed stereotype content for gay male subgroups explains overall ratings. *Social Cognition*, *23*, 161–181. <https://doi.org/10.1521/soco.23.2.161.65626>
- Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*(1), 155–159. <https://doi.org/10.1037/0033-2909.112.1.155>

- Cuddy, A. J. C., Fiske, S. T., & Glick, P. (2007). The BIAS map: Behaviors from intergroup affect and stereotypes. *Journal of Personality and Social Psychology, 92*, 631–648. <https://doi.org/10.1037/0022-3514.92.4.631>
- Cuddy, A. J. C., Fiske, S. T., Kwan, V. S. Y., Glick, P., Demoulin, S., Lexens, J.-P., . . . Ziegler, R. (2009). Stereotype content model across cultures: Towards universal similarities and some differences. *British Journal of Social Psychology, 48*, 1–33. <https://doi.org/10.1348/014466608X314935>
- Durante, F., Fiske, S. T., Gelfand, M. J., Crippa, F., Suttora, C., Stillwell, A., . . . Teymoori, A. (2017). Ambivalent stereotypes link to peace, conflict, and inequality across 38 nations. *Proceedings of the National Academy of Sciences of the United States of America, 114*, 669–674. <https://doi.org/10.1073/pnas.1611874114>
- Durante, F., Fiske, S. T., Kervyn, N., Cuddy, A. J., Akande, A. D., Adetoun, B. E., . . . Barlow, F. K. (2013). Nations' income inequality predicts ambivalence in stereotype content: How societies mind the gap. *British Journal of Social Psychology, 52*, 726–746. <https://doi.org/10.1111/bjso.12005>
- Eckes, T. (2002). Paternalistic and envious gender stereotypes: Testing predictions from the stereotype content model. *Sex Roles, 47*, 99–114. <https://doi.org/10.1023/A:1021020920715>
- Fiske, S. T. (2017). Prejudice in cultural contexts: Shared stereotypes (gender, age) versus variable stereotypes (race, ethnicity, religion). *Perspectives on Psychological Science, 12*, 791–799. <https://doi.org/10.1177/1745691617708204>
- Fiske, S. T., Cuddy, A. J. C., Glick, P., & Xu, J. (2002). A model of (often mixed) stereotype content: Competence and warmth respectively follow from perceived status and competition. *Journal of Personality and Social Psychology, 82*, 878–902. <https://doi.org/10.1037/0022-3514.82.6.878>
- Friehts, M.-T., Kotzur, P. F., Zöller, A.-K., Wagner, U., & Asbrock, F. (2020). *German published Stereotype Content Model scales are more often structurally invalid than they are valid: Analyses and implications*. Manuscript submitted for publication.
- Fröhlich, L., & Schulte, I. (2019). Warmth and competence stereotypes about immigrant groups in Germany. *PLoS ONE, 14*. <https://doi.org/10.1371/journal.pone.0223103>
- Gelfand, M. J., Jana, L. R., Nishii, L., Leslie, L. M., Lun, J., Lim, B. C., . . . Yamaguchi, S. (2011). Differences between tight and loose cultures: A 33-nation study. *Science, 332*, 1100–1104. <https://doi.org/10.1126/science.1197754>
- Graham, J. W., Taylor, B. J., Olchowski, A. E., & Cumsille, P. E. (2006). Planned missing data designs in psychological research. *Psychological Methods, 11*, 323–343. <https://doi.org/10.1037/1082-989X.11.4.323>
- Hodson, G., & Busseri, M. A. (2012). Bright minds and dark attitudes: Lower cognitive ability predicts greater prejudice through right-wing ideology and low intergroup contact. *Psychological Science, 23*, 187–195. <https://doi.org/10.1177/0956797611421206>
- Hofstede, G. (1980). *Culture's consequences: International differences in work-related values*. Beverly Hills, CA: Sage Publications.
- Janssens, H., Verkuyten, M., & Khan, A. (2015). Perceived social structural relations and group stereotypes: A test of the Stereotype Content Model in Malaysia. *Asian Journal of Social Psychology, 18*(1), 52–61. <https://doi.org/10.1111/ajsp.12077>
- Kervyn, N., Fiske, S., & Yzerbyt, V. (2015). Forecasting the primary dimension of social perception. *Social Psychology, 46*(1), 36–45. <https://doi.org/10.1027/1864-9335/a000219>
- Kline, R. B. (2010). *Principles and practise of structural equation modelling* (3rd ed.). New York, NY: Guilford.
- Koch, A., Imhoff, R., Dotsch, R., Unkelbach, C., & Alves, H. (2016). The ABC of stereotypes about groups: Agency/socioeconomic success, conservative-progressive beliefs, and communion. *Journal of Personality and Social Psychology, 110*, 675–709. <https://doi.org/10.1037/pspa0000046>
- Kotzur, P. F., Forsbach, N., & Wagner, U. (2017). Choose your words wisely: Stereotypes, emotions, and action tendencies toward fled people as a function of the group label. *Social Psychology, 48*, 226–241. <https://doi.org/10.1027/1864-9335/a000312>

- Kotzur, P. F., Friehs, M.-T., Asbrock, F., & Van Zalk, M. H. W. (2019). Stereotype content of refugees in Germany. *European Journal of Social Psychology, 49*, 1344–1358. <https://doi.org/10.1002/ejsp.2585>
- Kotzur, P. F., Schäfer, S. J., & Wagner, U. (2019). Meeting a nice asylum seeker: Intergroup contact changes stereotype content perceptions and associated emotional prejudices, and encourages solidarity-based collective action intentions. *British Journal of Social Psychology, 58*(3), 668–690. <https://doi.org/10.1111/bjso.12304>
- Lancee, B., Birkelund, G., Coenders, M., Di Stasio, V., Fernández Reino, M., Heath, A., . . . Zwier, D. (2019). *The GEMM study: A cross-national harmonized field experiment on labour market discrimination: Technical report*. Retrieved from <https://gemm2020.eu/wp-content/uploads/2019/02/GEMM-WP3-technical-report.pdf> (January, 2020)
- Lavrakas, P. J. (2008). Social desirability. In P. J. Lavrakas (Ed.), *Encyclopaedia of survey research methods* (pp. 411–412). Thousand Oaks, CA: Sage Publications. <https://doi.org/10.4135/9781412963947>
- Leach, C. W., Ellemers, N., & Barreto, M. (2007). Group virtue: The importance of morality (vs. competence and sociability) in the positive evaluation of in-groups. *Journal of Personality and Social Psychology, 93*, 234–249. <https://doi.org/10.1037/0022-3514.93.2.234>
- Muthén, L. K., & Muthén, B. O. (1998–2017). *Mplus user's guide* (8th ed.). Los Angeles, CA: Muthén & Muthén.
- Muthén, L. K., & Muthén, B. O. (2002). How to use a Monte Carlo study to decide on sample size and determine power. *Structural Equation Modeling, 9*, 599–620. https://doi.org/10.1207/S15328007SEM0904_8
- Ross, L., Greene, D., & House, P. (1977). The “false consensus effect”: An egocentric bias in social perception and attribution processes. *Journal of Experimental Social Psychology, 13*, 279–301. [https://doi.org/10.1016/0022-1031\(77\)90049-X](https://doi.org/10.1016/0022-1031(77)90049-X)
- Sadler, M. S., Meagor, E. L., & Kaye, K. E. (2012). Stereotypes of mental disorders differ in competence and warmth. *Social Science & Medicine, 74*, 915–922. <https://doi.org/10.1016/j.socscimed.2011.12.019>
- Schermelleh-Engel, K., Moosbrugger, H., & Müller, H. (2003). Evaluating the fit of structural equation models: Tests of significance and descriptive goodness-of-fit measures. *Methods of Psychological Research Online, 8*, 23–74.
- Schwartz, S. H. (2011). Values: Cultural and individual. In F. J. R. van de Vijver, A. Chasiotis & S. M. Breugelmans (Eds.), *Fundamental questions in cross-cultural psychology* (pp. 463–493). Cambridge, UK: Cambridge University Press.
- Stanciu, A., Cohrs, J. C., Hanke, K., & Gavreliuk, A. (2017). Within-culture variation in the content of stereotypes: Application and development of the stereotype content model in an Eastern European culture. *The Journal of Social Psychology, 157*, 611–628. <https://doi.org/10.1080/00224545.2016.1262812>
- Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods, 3*(1), 4–70. <https://doi.org/10.1177/109442810031002>
- Veit, S., & Yemane, R. (2018). *The ADIS study: A large-scale correspondence test on labor market discrimination in Germany – Technical Report*. WZB Discussion Paper, SP IV 2018–103. Berlin, Germany: WZB. Retrieved from <https://bibliothek.wzb.eu/pdf/2018/vi18-103.pdf> (January, 2020).
- Veit, S., & Yemane, R. (2020). *Judging without knowing: how people evaluate others based on phenotype and country of origin (Technical Report)*. WZB Discussion Paper, SP VI 2020-101, Berlin, Germany: WZB.
- Webster, R. J., Burns, M. D., Pickering, M., & Saucier, D. A. (2014). The suppression and justification of prejudice as a function of political orientation. *European Journal of Personality, 28*(1), 44–59. <https://doi.org/10.1002/per.1896>

Supporting Information

The following supporting information may be found in the online edition of the article:

Table S1 Study 2: Model Fit Indices of Measurement Models for each Social Group and Condition Separately, Using All Items.

Table S2 Study 2: Model Fit Indices of Measurement Models for each Social Group and Condition Separately, Excluding the Item Warm.

Table S3 Study 2: Model Fit Indices of Measurement Models for each Social Group and Condition Separately, Excluding the Items Warm and Confident.

Table S4 Study 2: Model Fit Indices of Measurement Models for each Social Group and Condition Separately, Excluding the Items Warm, Confident, and Competitive.

Table S5 Study 2: Count of Items that Emerged as Problematic in Model Fit Indices ($\Delta\chi^2 \geq 10$).

Table S6 Study 3: Model Fit Indices of Measurement Models for Each Group and Condition Separately, Using all Items.

Table S7 Study 3: Model Fit Indices of Measurement Models for Each Group and Condition Separately, Excluding the Item cold–warm.

Table S8 Study 3: Count of Items that Emerged as Problematic in Model Fit Indices ($\Delta\chi^2 \geq 10$).