Empirical Research Paper

The Vicissitudes of Paternalistic Stereotypes: Is Warmth Toward Socially Protected Groups Authentic?

Personality and Social Psychology Bulletin
1–15

DOI: 10.1177/01461672251378535
journals.sagepub.com/home/pspb
Sage

Vincent Yzerbyt[1], Odile Rohmer[2], and Laurent Cambon[3]

Abstract

The present endeavor examines whether paternalistic stereotypes, associating some low-status groups with low competence but high warmth, are always genuine. We propose that not all individuals are sincere in their evaluations as their judgments of warmth are likely driven by normative concerns. Study 1 ($N = 201$) confirmed the relationship between paternalistic stereotypes and perceived normative protection. Three additional studies manipulated participants' beliefs about whether their true opinion could be uncovered or was socially expected. Specifically, individuals with high external motivation to respond without prejudice rated socially protected groups more harshly, especially on warmth, when they faced a bogus pipeline (Study 2, $N = 160$) or when norms emphasized honesty (Studies 3a and 3b, $Ns = 137$ and 247). These findings show the greater versatility of warmth compared to competence evaluations and provide new insights into why members of protected groups may continue to experience discrimination.

Keywords

stereotypes, social evaluation, Big Two, normative protection, motivation to respond without prejudice

Received February 10, 2025; revision accepted August 28, 2025

Many minority groups continue to face discrimination even if they are frequently described with a blend of negative and positive traits and would seem to be socially protected (Czopp et al., 2015; Kay et al., 2013). For example, people with disabilities are seen as poor in intellectual or physical capacities but considered as nice and admirable individuals (Granjon et al., 2023). That prejudice against them would seem to persist raises an intriguing question: do respondents generally express their genuine views about these groups, or do social expectations influence their responses? In other words, when respondents attribute positive traits to such groups, are they revealing their honest attitudes? The possibility that some respondents might be reluctant to disclose their actual negative views of these socially protected groups could explain why there often remains a gap between individuals' words and actions (Ilmarinen et al., 2022; Rohmer et al., 2024). This set of studies seeks to explore this issue further, investigating whether certain motivations drive people to appear less prejudiced than they truly are.

Core Dimensions of Social Perception: From the Big Two to Paternalistic Stereotypes

Building upon a rich array of theoretical and empirical efforts such as the Stereotype Content Model (SCM; Fiske et al., 2002) and the BIAS Map (Cuddy et al., 2008), current understanding suggests that social perception is guided by two core dimensions, often called the Big Two, which shape how we evaluate groups, individuals, and ourselves (for recent reviews, see Abele et al., 2021; Yzerbyt et al., 2025). The "horizontal" dimension, also named warmth or communion, concerns morality and sociability. This dimension gauges whether social targets are trustworthy and cooperative or whether caution is needed. The "vertical" dimension, also named competence or agency, relates to the resources, skills, power, and prestige of individuals or groups. This dimension assesses whether targets can achieve their goals or are likely to face obstacles (Abele et al., 2021; Abele & Wojciszke, 2014; Fiske et al., 2007). Together, the Big Two form a two-dimensional space that comprises four quadrants combining low and high levels of the horizontal and vertical dimensions.

[1]Catholic University of Louvain, Belgium
[2]University of Strasbourg, France
[3]University of Grenoble-Alpes, France

Corresponding Author:
Vincent Yzerbyt, Institut de Recherche en Sciences Psychologiques, Université catholique de Louvain, Place du Cardinal Mercier, Louvain-la-Neuve 1348, Belgium.
Email: vincent.yzerbyt@uclouvain.be

An intriguing distinction arises among social groups low on the vertical dimension. These groups are often seen as lacking resources, power, and competence. But while those low in vertical/low in horizontal (e.g., drug addicts, child abusers, homeless) are devalued entirely, those low in vertical/high in horizontal (e.g., the elderly, housewives, people with disabilities) are characterized by a so-called paternalistic stereotype, that is, they are seen as incompetent but warm, often eliciting sympathy (Cuddy et al., 2008).

The Impact of the Normative Context

The fact that some groups trigger paternalistic judgments aligns well with findings that certain groups are perceived to be more socially protected than others. In a seminal contribution, Crandall and colleagues (2002) showed that the public expression of prejudice toward 105 social groups was strongly correlated with the social approval of such expressions and that this was particularly true for low vertical/high horizontal groups (Jeffries et al., 2012). The authors argued, however, that the suppression of prejudice is not solely driven by internal values like egalitarianism but is also likely to be motivated by a desire to conform to perceived social norms about the appropriateness of expressing prejudice. Given the prevalence of protective norms for various social groups, a key question thus arises: do the paternalistic stereotypes associated with these groups always reflect perceivers' genuine opinions, or are some responses primarily shaped by social norms?

Several empirical studies suggest that respondents do not always uncritically endorse the stereotypical opinion about protected groups as being high on the horizontal dimension. For instance, previous studies found mixed stereotypes of people with disabilities on self-report measures, whereas indirect measures (via IAT or LDT) revealed non-mixed stereotypes (low competence/low warmth; Clément-Guillotin et al., 2018; Rohmer & Louvet, 2012). Along similar lines, Schmitz and Yzerbyt (2022) found that members of a high-status minimal group judged a low-status group as less competent but warmer than themselves on direct, self-report, measures. In contrast, on an indirect measure, i.e., the Brief-IAT, these same participants denied the greater warmth of the low-status group. As other authors have suggested (Bergsieker et al., 2012; Rohmer & Louvet, 2018), members of high-status groups may express positive views regarding the warmth of the incompetent group to conform to normative concerns and relationship considerations, thereby displaying magnanimity toward specific low-status groups (see also, Cambon & Yzerbyt, 2018).

Next to the demonstration that normative contexts shape people's evaluations, a key takeaway from these studies is their focus on the horizontal dimension, reflecting findings that this dimension is more subjective and therefore more "negotiable" than the vertical dimension. In the early presentations of the SCM, Fiske and colleagues (2002) noted that structural indicators and psychological evaluations correlate more strongly along the vertical than the horizontal dimension. Similarly, Koch and colleagues (2020) suggested that agency, a facet of the vertical dimension, is the primary and most agreed-upon aspect observed in social settings. Also, in research on the Dimensional Compensation Model (Yzerbyt, 2018), when participants were asked to rate the objectivity of both dimensions, results showed that horizontal judgments allowed for greater subjective freedom (Yzerbyt & Cambon, 2017).

Overall, thus, positive evaluations on the horizontal can be particularly useful when aiming to comply with non-discrimination standards. This distinction between the horizontal and the vertical dimensions, along with the idea that the horizontal is probably crucial in revealing the normative protection that people want to grant to a group, allows specifying Crandall et al.'s approach in new and important ways.

Interindividual Differences in Normative Concerns

Building on the insight that warmth ascribed to groups with low vertical status may partly stem from people's tendency to conform to social norms, we wanted to go beyond earlier efforts (Rohmer & Louvet, 2012; Schmitz & Yzerbyt, 2022) and propose that not everyone is equally influenced by normative pressures. Indeed, whereas for some perceivers, norms align with their true beliefs, others are likely to adjust their evaluations when this is possible and desirable, particularly along the horizontal dimension. Drawing from Dunton and Fazio (1997) and Plant and Devine (1998), we propose that individuals with high external motivation to respond without prejudice may view protected groups less genuinely positively than those who hold internalized positive beliefs. When a situation discourages people from hiding their true beliefs, high externally motivated participants may reveal more negative attitudes toward protected groups.

Interestingly, research highlights an even broader personality trait related to people's sensitivity to social norms: self-monitoring (Snyder, 1974). The self-monitoring scale measures how much individuals adjust their behavior to manage their public image. High self-monitors are particularly responsive to social expectations, often modifying their behavior to align with the attitudes of those around them to appear socially adaptive. In contrast, low self-monitors tend to act according to their internal beliefs, staying consistent regardless of social pressures (Klein et al., 2004).

This reasoning suggests that individuals with high external motivation to respond without prejudice or high self-monitoring tendencies often express favorable views of socially protected groups, mainly due to what they perceive to be the social expectations. In a context that encourages honesty, these individuals may shift away from these positive evaluations and express less favorable judgments, especially on the horizontal dimension. This is because, in a

setting that incites sincerity, their evaluations are less influenced by the norms that typically promote non-discriminatory responses. In summary, while protected groups are stereotyped in a paternalistic way, that is, as low in competence (vertical dimension) but high in warmth (horizontal dimension), this pattern may weaken among those who usually report such judgments primarily due to social pressure rather than personal conviction. These individuals may show less positive evaluations when the context leads them to downplay the norm of non-discrimination.

The Present Studies

The current research sought to examine whether individuals who are highly sensitive to social norms will deviate from the usual mixed stereotype of socially protected groups (low in competence but high in warmth) and disclose more negative views on warmth when the context pushes them to share their true opinions. We conducted four studies to tackle this hypothesis.

Study 1 aimed to show that it is essentially those groups perceived as low on competence but high warmth that people consider socially protected. In the next three studies, we sought to create different normative contexts such that individual differences between low and high externally motivated participants would show in their evaluations of protected groups, particularly on warmth. Study 2 employed the bogus pipeline procedure to reduce social desirability bias and potentially reveal less favorable judgments of protected groups. We expected that participants with high external motivation to respond without prejudice would be most sensitive to this change in context, displaying more candid and, indeed, harsher evaluations in the presence of the bogus pipeline. Studies 3a and 3b refined this approach by directly tackling the nature of the contextual norms. To ease social pressure to give non-discriminatory responses, we informed some participants that their peers greatly valued honesty in expressing judgments. Study 3b expanded on Study 3a by including a larger set of groups and traits and measuring self-monitoring. Across Studies 2–3b, we hypothesized that participants typically driven by external pressures for non-discriminatory evaluations would provide more critical horizontal evaluations of protected groups when usual social constraints were relaxed. Although we also measured internal motivation to respond without prejudice in all these studies and considered this factor along with external motivation, we had no specific hypotheses regarding its impact on participants' judgments.

Study 1

Study 1's primary ambition was to select a variety of groups in terms of their position in the four quadrants defined by the SCM and to check whether the paternalistic stereotypes that characterize certain groups, i.e., those low on competence (the vertical dimension) but high on warmth (the horizontal dimension), go hand in hand with the perception that these groups are socially protected. Participants indicated the degree of normative protection that they perceived associated with each group, in addition to each group's perceived standing on the competence and warmth. A second goal was to adapt the internal and external motivation to respond without prejudice scale in French to be able to use this individual difference in subsequent studies.

Method

Participants. Participants were psychology students at a large French university and were recruited for partial course credit. They received information on the purpose of the study, learned that the study was anonymous, and that they were free to stop at any time. We relied on the same steps for all remaining studies. After discarding the data from five participants (three participants in list 1 condition and two participants in list 2) who failed to complete the questionnaire, the final sample comprised 201 participants (42 % female). Ages ranged between 17 and 54 years (M = 22.16; SD = 5.96).

Procedure and Design. Participants completed the questionnaire online. Given the number of measures and to reduce the burden in filling in the questionnaire, participants randomly received one of two lists of groups ($N_{list\ 1} = 106$; $N_{list\ 2} = 94$). The first list comprised 23 groups and the second 24. Participants first indicated the normative protection they perceived as being associated with each group. Next, participants rated each group's warmth and competence. They then reported their internal versus external motivation to respond without prejudice. Finally, they completed demographic questions, including age and gender.

Materials. A pretest (N = 41) allowed to secure an initial set of groups. Participants had to list between 6 and 12 types of people that today's society considers as groups. Because participants spontaneously named fewer groups than expected, we added a series of groups taken from Crandall et al.'s (2002) and Jeffries et al.'s (2012) studies. The final list included national groups (e.g., French, Brazilians), occupational groups (e.g., Architect, Barman), ethnic groups (e.g., Arab, Asiatic), stigmatized groups (e.g., Migrant, Homeless), and more general social categories (e.g., Men, Women). We checked the level of competence and warmth of each group from previous studies to secure a priori approximately equal numbers of groups for each quadrant of the two-dimensional space created by the Big Two (Abele et al., 2021).

Measures[1]

Perceived Normative Protection. Participants rated each group on a 7-point scale ranging from 1 (totally socially

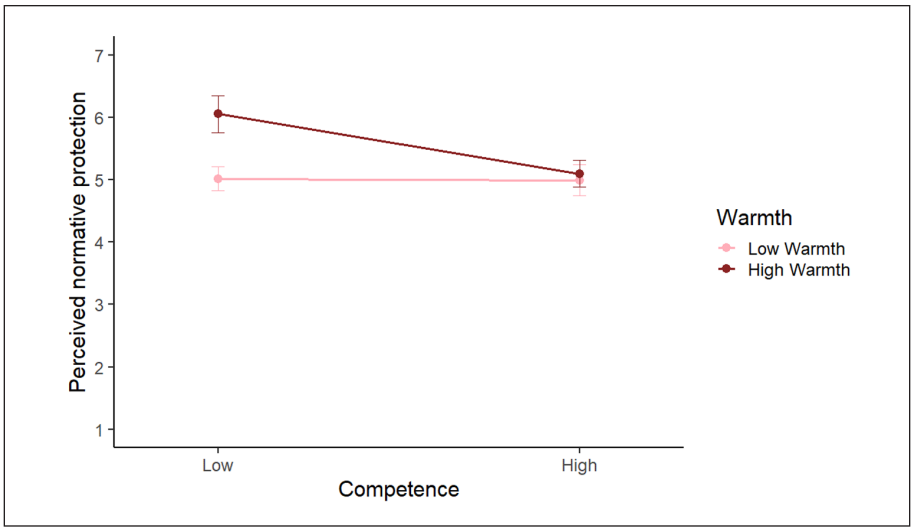

Figure 1. Perceived Normative Protection of the Groups as a Function of Warmth and Competence.

unacceptable to express a negative attitude about this group) to 7 (totally socially acceptable to express a negative attitude about this group). In full accordance with Crandall and colleagues' (2002) original measure, the instructions given to participants made clear that the ratings were about their perceptions of the predominant social norm and not their own attitudes.

Stereotypes. The items pertaining to warmth, and competence of each group were preceded by the same stem reading: "*In today's French society, _____ are perceived as a group . . .*" followed by: "of *warm persons,"* and "*of friendly persons"* (warmth scales), " *of competent persons"* and "*of capable persons"* (competence scales). Participants answered all questions on 5-point scales ranging from 1 (*not at all*) to 5 (*totally*).

Internal and External Motivation to Respond Without Prejudice. We translated in French Plant and Devine's (1998) scale with five internal and five external items assessed on a 7-point scale ranging from 1 (*I totally disagree with this affirmation*) to 7 (*I totally agree with this affirmation*). We adapted the items to measure the motivation to respond without prejudice toward stigmatized groups in general (see Study instructions and materials; https://osf.io/y6un5/files/osfstorage?view_only=None)

Results

Perceived Normative Protection. To examine the relation between perceived normative protection and group stereotypes on the fundamental dimensions, we first calculated the participant-level correlations between our two competence items (competent, capable) and between our two warmth items (warm, friendly) for each group. Across the 47 groups, these correlations averaged to $r = .84$ ($SD = .09$) for the former and $r = .88$ ($SD = .08$) for the latter. We then combined the relevant items across participants to form one warmth score and one competence score for each group (see Table S1 in Supplementary Materials).

Next, we regressed the perceived normative protection of the groups on their (centered) competence and warmth scores along with the interaction between these two scores (see Figure 1).[2] Whereas competence predicted protection negatively, $b = –0.35$, $t(43) = –3.92$, $p < .001$, $\eta_p^2 = .26$, warmth showed a positive relation, $b = 0.48$, $t(43) = 4.78$, $p < .001$, $\eta_p^2 = .35$. Finally, the interaction term also proved significant, $b = –0.57$, $t(43) = –4.03$, $p < .001$, $\eta_p^2 = .27$. Further examining the simple effects at +1/–1 *SD* revealed that participants perceived very different levels of normative protection for high warmth groups, with low competence groups seen as enjoying more normative protection ($M = 6.06$, $SE = 0.15$) than high competence groups ($M = 5.10$, $SE = 0.11$), $b = –0.69$, $t(43) = –4.91$, $p < .001$, $\eta_p^2 = .36$. No such difference emerged for low warmth groups, with low and high competence groups seen as enjoying similar levels of normative protection ($M = 5.02$, $SE = 0.09$, and $M = 5.00$, $SE = 0.12$), $b = –0.02$, $t(43) = –0.16$, ns. Conversely, warmth hardly affected how participants perceived normative protection targeting high competence groups, $b = 0.09$, $t(43) = 0.62$, ns. In sharp contrast, low competent groups came across as enjoying much more normative protection when participants also saw them as warm as opposed to cold, $b = 0.88$, $t(43) = 6.15$, $p < .001$, $\eta_p^2 = .47$.

Internal and External Motivation to Respond Without Prejudice. We performed exploratory factor analyses with Oblimin rotation on the 10 items of the internal and external motivation to respond without prejudice scale (IEM). Two factors explained a total of 64% of the variance and correlated negatively, $r = –.24$.[3] As expected, one factor comprised the

external motivation items, whereas the other included the internal motivation ones (all factor loadings > .56 on their respective factor). This allowed creating two scales with very good reliability ($\alpha_{external} = .86$ and $\alpha_{internal} = .83$)

Discussion

Study 1 achieved two important goals. First, we confirmed that perceived normative protection was related to the degree to which perceivers associated groups with paternalistic stereotypes, that is, as being simultaneously warm and not competent. This finding sets the stage for a deeper examination of the role of differences in perceivers' sensitivity to normative contexts on the expression of paternalistic stereotypes. Another dividend of Study 1 is that our French version of the IEM scale proved quite satisfactory. This enables us to use this scale developed to measure motivation not to discriminate against Blacks in an American context to tap external and internal motivation not to discriminate against social minorities in contemporary France.

In Study 2, we capitalized on Study 1 to select a limited number of groups perceived to be protected. We wanted to examine how participants would evaluate these groups in the absence of the usual normative pressures that are particularly strong for these specific groups.

Study 2

In Study 2, we opted for a bogus pipeline procedure (Jones & Sigall, 1971). The bogus pipeline is a method where participants are falsely made to believe that some device (a machine or some other technical or methodological procedure) can accurately detect lies or measure their true attitudes or feelings (e.g., a fake lie detector). This belief typically leads participants to provide more truthful responses and has been used repeatedly to tap attitudes and opinions when social desirability may come in the way of accessing a sincere answer on the part of the respondents (Tourangeau et al., 1997).

To be sure, it is hardly surprising that groups perceived to be socially protected are also those seen as lacking in the vertical dimension. In other words, a low level of competence is an objective rendition of the social hierarchy (Yzerbyt et al., 2022). Because such a stereotypical evaluation conveys a negative image, one may rightly consider it to be rather sincere. In contrast, because people also judge these same groups as rather warm, i.e., a positive stereotype, a legitimate question is whether this evaluation is genuine. Indeed, some people may disagree with this stereotype while being concerned about the acceptability of their evaluations.

Building on this reasoning, we hypothesized that, compared to control participants, the confrontation with a bogus pipeline would cause participants to become more truthful and be less positive in their evaluations, particularly on the warmth dimension. Moreover, because research by Plant and Devine (1998, 2001) showed that only those who are internally motivated have a genuine desire for a more egalitarian society, we conjectured that high externally motivated individuals would be the ones most sensitive to the bogus pipeline procedure. Whereas they should value minorities on warmth when this poses no problem, that is, in the control condition, and express paternalistic stereotypes, they should convey more negative judgments on warmth in the presence of the bogus pipeline. This negative shift in warmth should be absent or at least less marked among participants scoring low on the external motivation to respond without prejudice.

To sum up, we predicted a three-way interaction such that, compared to participants low on external motivation to respond without prejudice, those high on external motivation to respond without prejudice would express harsher judgments on warmth (and less so on competence) when confronted to the bogus pipeline.

Method

Design and Participants. Relying on the mean effect size ($d = .41$) of the bogus pipeline procedure reported in Roese and Jamieson's (1993) meta-analysis, we turned to the PANGEA Webapp (https://jakewestfall.shinyapps.io/pangea/).[4] Our design was a 2 (condition: bogus pipeline vs. control) × (centered) external motivation to respond without prejudice × (centered) internal motivation to respond without prejudice × 2 (dimension: warmth vs. competence) with the first three variables varying between participants and the last within them. We considered the power for our three-way interaction involving condition by external motivation to respond without prejudice by dimension. This analysis suggested that we needed a total of 136 participants to achieve 80% power. We increased this number by some 20% to take an overestimation of the effect size as well as potential data loss into account.[5] After discarding the data from three participants (two in the control condition and one in the bogus pipeline condition) because they either had not completed the questionnaire ($n = 2$) or had no variance in their results ($n = 1$), the final sample comprised 160 participants (81 % female). Ages ranged between 18 and 67 years ($M = 22.53$; $SD = 8.86$).

Measures[6]

Internal and External Motivation to Respond Without Prejudice. Participants completed the same 10-item IEM scale as in Study 1.

Warmth and Competence Judgments. Participants rated elderly people and people with disabilities on two warmth (warm, friendly) and two competence (competent and capable) traits on scales ranging from 1 (*not at all*) to 7 (*totally*). We selected these two groups because Study 1 showed that they were associated with very high perceived normative protection scores, along with the strongest difference between warmth and competence. The correlations between

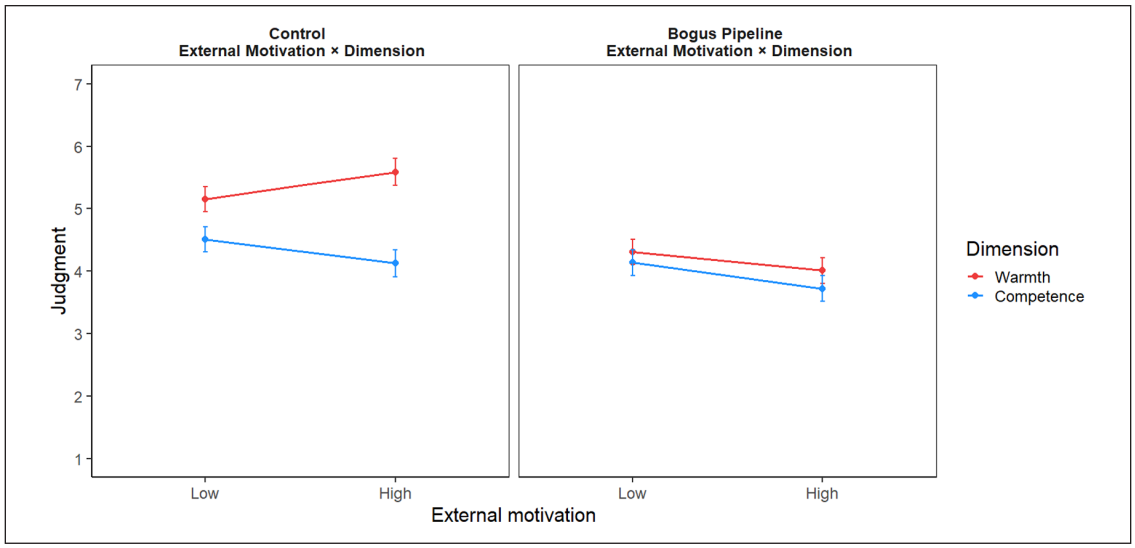

Figure 2. Judgments of Warmth and Competence as a Function of Condition and External Motivation to Respond Without Prejudice.

the two items, ranging between .71 and .83, allowed us to compute a global score for each dimension and each group. We then collapsed the scores over the groups.

Manipulation Checks. Participants answered a series of questions pertaining to their opinion about the Lexical Decision Task (LDT, see below). The questions read "Do you think that reaction time is a good way to measure people's attitude?," "Do you think that the LDT is a reliable measure of attitude?," "Do you think that LDT measured reliably your attitude toward the two social groups?," "Were you able to control the speed of your answers to trump the measure? (reversed)," and "Do you think that LDT could be used in judicial context to detect lies" on a scale ranging from 1 (= *not at all*) to 7 (= *totally*). As the five manipulation check items showed very good internal consistency ($\alpha = .74$), we averaged them in a single score.

Procedure. The experiment comprised two sessions. In the first, participants completed the IEM scale. One week later, we contacted them again for an apparently unrelated study, allegedly testing a new way to measure attitudes. We relied on a LDT that was presented differently in the two conditions (for a similar procedure, see Nier, 2005). Specifically, we informed control participants that the experiment aimed to compare two different measures of attitudes. They were first given the LDT, measuring attitudes toward both elderly people and people with disabilities. They then completed a short online questionnaire measuring their views about these two groups on warmth and competence scales. As to bogus pipeline participants, they learned that the experiment aimed to test the generalizability of a new reliable measure of the "true" attitude. They then went on to the same steps as control participants except that we presented the LDT as a new version of a lie detector, that they then had to report their impressions of the two groups on the rating scales as they thought the bogus pipeline had detected them, and that they concluded their session by responding to questions about their perception of the LDT. We randomly assigned participants to one of these two conditions. Finally, participants were debriefed, thanked, and dismissed.

Results

Manipulation Check. A one-way ANOVA on the manipulation check score confirmed the success of our manipulation, $F(1, 158) = 11.66$, $p < .001$, $\eta_p^2 = .069$. Participants reported being more convinced by the LDT as a "lie detector" in the bogus pipeline condition ($M = 4.93$, $SD = 1.23$) than in the control condition ($M = 4.36$, $SD = 0.86$).

Warmth and Competence Judgments. Next, we submitted the warmth and competence scores of the two groups to a mixed-model analysis using condition (with control condition = –1/2, and bogus pipeline condition = 1/2), dimension (warmth = –1/2 and competence = 1/2), (centered) external and (centered) internal motivation to respond without prejudice along with their interactions as our fixed effects and participants as our random effect.[7]

For the sake of brevity, we only report the result pertaining to our critical interaction, namely condition by dimension by external motivation to respond without prejudice. The full results can be found in the Supplementary Materials.

As expected, the condition by dimension by external motivation interaction proved significant, $F(1, 470) = 16.46$, $p < .001$, $= .03$ (see Supplemental Table S2). We decomposed this interaction by dimension (see Figure 2). Turning to warmth first, we found a very significant

condition effect, $F(1, 272) = 98.70$, $p < .001$, $\eta_p^2 = .27$. More importantly, and in line with our hypothesis, there was a condition by external motivation interaction, $F(1, 272) = 19.18$, $p < .001$, $\eta_p^2 = .07$. Whereas low externally motivated participants rated the groups as being colder in the bogus pipeline than in the control condition ($M = 5.05$, $SE = 0.12$, and 4.38, $SE = 0.12$), $t(272) = –3.92$, $p < .001$, $\eta_p^2 = .05$, this pattern was even more pronounced for the high externally motivated ($M = 5.69$, $SE = 0.13$, and 3.94, $SE = 0.12$), $t(272) = –9.88$, $p < .001$, $\eta_p^2 = .26$. Conversely, whereas the control condition led high externally motivated to rate the groups warmer than low externally motivated did, $t(272) = 3.62$, $p < .001$, $\eta_p^2 = .05$, the opposite pattern materialized in the bogus pipeline condition, $t(272) = –2.56$, $p < .01$, $\eta_p^2 = .02$. As to the analysis for the competence ratings, we found a main effect of condition, $F(1, 272) = 10.20$, $p < .002$, $\eta_p^2 = .04$. Participants saw the groups as more competent in the control ($M = 4.32$, $SE = 0.09$) than in the bogus pipeline condition ($M = 3.93$, $SE = 0.09$). Also, there was a main effect of external motivation to respond without prejudice, $F(1, 272) = 23.09$, $p < .001$, $\eta_p^2 = .08$, in that low externally motivated gave the groups higher competence ratings ($M = 4.42$, $SE = 0.09$) than high externally motivated did ($M = 3.83$, $SE = 0.09$).

Of note, none of the three-way interactions involving internal motivation to respond without prejudice proved significant, all $Fs < 1$, *ns*. Also, the analysis revealed no significant four-way interaction, $F(1, 470) = 2.03$, $p < .16$.

Discussion

The results of Study 2 prove most encouraging in that we found the predicted three-way interaction involving condition, external motivation to respond without prejudice, and dimension. Whereas high externally motivated participants complied with the paternalistic stereotype about the groups when they were in the control condition (and indeed more so than low externally motivated), their evaluation deteriorated in the bogus pipeline condition, especially with respect to the evaluation of warmth. No such pattern emerged with respect to the internal motivation to respond without prejudice.

Although the instructions associated with the bogus pipeline condition targeted the causal factor we had in mind, a critical reader might argue that our manipulation may have been too blatant. Because the context made it explicit that people might lie when conveying their judgments, participants may have been cognizant of the intentions of the experimenter and reacted accordingly. In other words, bogus pipeline participants may have understood that they were expected to depart from their spontaneous impression to propose a different, possibly harsher, evaluation of the targets. One reason to doubt this explanation resides in the fact that bogus pipeline participants were quite selective in changing their response pattern and modified only the warmth but not the competence ratings. Although the literature has repeatedly confirmed the efficiency of the bogus pipeline manipulation (Roese & Jamieson, 1993), we deemed it important to replicate the pattern observed in Study 2 with an entirely different procedure, one that closely corresponds to real-life contexts. Finally, we also wanted to consider additional protected groups, allowing us to take group variability better into account.

Studies 3a and 3b

In Study 3a and 3b, we aimed to replicate the pattern observed in Study 2 by capitalizing on the well-established influence of descriptive norms on behavior (Cialdini et al., 1991; Stangor et al., 2001). We created two situations in which participants learned that the prevalent norms promoted either non-discrimination or honesty. An additional goal was to increase the number of groups (five in Study 3a and seven in Study 3b) and the number of traits (Study 3b). Finally, we also wanted to explore the role of people's sensitivity to social norms beyond the motivation to respond without prejudice by examining participants' self-monitoring (Study 3b). Concretely, participants first completed the IEM scale and, in Study 3b, also the self-monitoring scale. Next, we asked their views about various low competence/high warmth groups subject to normative protection and did so under different normative conditions. In the "honesty" condition, participants rated the groups after seeing the results of a survey suggesting that most people consider it important to express their honest views about other groups rather than hiding them for the sake of not discriminating against others. In the "non-discrimination" condition, the message of the survey was instead that most people saw it critical to avoid discriminating social groups and that political correctness constitutes a relevant strategy to avoid offending anyone. Next to these two conditions, Study 3a also included a third "control" condition in which participants evaluated the groups without prior normative information. Importantly, we preregistered Study 3b: https://osf.io/khq8b?view_only=83e59d844c704660be1baf3f24be25ba

Again, we hypothesized that the influence of social norms would be most effective among high externally motivated participants. That is, we hoped that participants who are preoccupied with their social image would prove sensitive to the prevalent norms. Compared to low externally motivated participants, high externally motivated should thus depart from the dominant paternalistic stereotypes about protected groups when they find themselves confronted with a social norm that promotes honesty and condemns political correctness. In this situation, we should expect harsher warmth judgments than in the non-discrimination norm or control conditions. Because of the prevalence of the normative protection of our target groups, we did not expect the latter two conditions to differ from each other.

Method

Design and Participants. Study 3a relied on a 3 (norm: non-discrimination *vs.* honesty *vs* control) × (centered) external motivation to respond without prejudice × (centered) internal motivation to respond without prejudice × 2 (dimension: competence vs. warmth) × 5 (target group) design with the first three variables varying between participants and the others within them. Relying on Study 2 which found a $\eta_p^2 = .03$ for the predicted three-way interaction, PANGEA suggested a total of 120 participants to achieve 82% power. To account for data loss and the potential overestimation of the effect size, we planned on contacting 140 participants. After discarding the data from six participants (one in the control condition, three in the non-discrimination condition, and two in the honesty condition) either because of many missing data (four) or no variance in their answers (two), the final sample consisted of 136 psychology students ($N_{\text{females}} = 105$, $N_{\text{males}} = 27$, $N_{\text{other}} = 4$) recruited for partial course credit. Ages ranged between 18 and 49 years ($M = 21.83$; $SD = 5.54$).[8]

In contrast to Study 3a, Study 3b was not conducted in one session but consisted of two phases, one for the measurement of our individual difference measures and one for the judgments of the groups. Notwithstanding this difference, Study 3b relied on the same design as Study 3a, except that we considered participants, target groups, and traits as random factors. Under these specifications and given the effect size found in Study 3a ($\eta_p^2 = .04$ for the three-way interaction, after dropping the control condition), the Pangea Webapp suggested a total of 192 participants to achieve 81% power. Considering data loss, possible overestimation of the effect size, but also, and importantly, an expected loss of roughly 50% of the participants between the two phases of the study, we planned 400 participants in Phase 1. The final sample comprised 247 psychology students ($N_{\text{females}} = 206$, $N_{\text{males}} = 36$, $N_{\text{other}} = 5$) from two large French universities. Ages ranged between 18 and 36 years ($M = 19.54$; $SD = 2.10$). We discarded eight participants (five in the non-discrimination condition and three in the honesty condition) because they did not report or incorrectly reported the code allowing us to link the answers of both phases.

Measures

Manipulation Check. We checked the success of the norm manipulation by asking participants in the honesty and non-discrimination conditions to report their agreement with four statements (e.g., "It is better to hide one's negative attitude toward stigmatized groups because it is important not to hurt them") on a 7-point scale going from 1 (*totally disagree*) to 7 (*totally agree*). Higher scores indicate stronger agreement with the non-discrimination norm. These scales showed very good internal consistency ($\alpha = .89$ for both studies), allowing us to compute a non-discrimination norm agreement score. Study 3b included two additional items on 7-point rating scales that checked for participants' specific memory of the survey results, $r = .80$. Higher scores indicated that participants thought the survey findings evidenced a preference for the non-discrimination norm as opposed to the honesty norm.

Internal and External Motivation to Respond Without Prejudice. We used the same IEM scale as in Studies 1 and 2.

Self-Monitoring Scale. In Study 3b, we used a French translation of Snyder's (1974) Self-Monitoring scale (Gana & Brechenmacher, 2001). Participants answered on a 7-point scale going from 1 (*totally disagree*) to 7 (*totally agree*). The good internal reliability ($\alpha = .79$) allowed us to compute a global self-monitoring score, which we centered. Higher scores indicated higher levels of self-monitoring.

Dependent Variables. Participants rated five groups on two competence traits (competent, capable) and two warmth traits (warm, friendly) in Study 3a. Scales ranged from 1 (= *not at all*) to 5 (= *totally*). The average of the correlations between the two items pertaining to each dimension reached .69 (ranging between .43 and .84), allowing us to compute a global score for each dimension and each group. For Study 3b, using similar scales, we relied on six competence traits (ambitious, self-assured, determined, competent, capable, intelligent) and six warmth traits (friendly, warm, likable, sincere, trustworthy, honest).

Procedure and Design. Students received the invitation to participate and a link to access the online questionnaires. During Phase 1, participants completed the IEM (and the self-monitoring scales for Study 3b). Immediately after this first phase (Study 3a) or one week later (Study 3b), they were asked to complete Phase 2, which consisted of an apparently unrelated set of two studies. The first study allegedly concerned political correctness and started with the presentation of the results of a survey on political correctness versus honesty. Participants were to give their opinion about the survey findings on a series of statements that served as a manipulation check. The second study focused on people's views about social groups, with questions assessing the competence and warmth of five (Study 3a) or seven (Study 3b) groups. To select the groups, we relied on Study 1 and ranked the groups according to the difference between warmth and competence from Study 1. We then made a random selection among those low competence/high warmth groups that were in the upper quarter of the list, namely obese people, nomad people, housewives, farmers, and Black (Study 3a), along with people with disabilities and the elderly (Study 3b). Finally, participants were debriefed, thanked, and dismissed.

Results

Manipulation Checks. A one-way ANOVA using the two norm conditions as our between-participants factor confirmed the success of our norm manipulation, $F_{Study3a}(1, 94) = 37.04$, $p < .001$, $\eta_p^2 = .28$. $F_{Study3b}$

Table 1. Means (Standard Errors) of Warmth and Competence as a Function of Condition and External Motivation to Respond Without Prejudice (Study 3a).

	Dimension					
	Warmth			Competence		
Condition	Non-discr.	Control	Honesty	Non-discr.	Control	Honesty
External motivation						
Low	3.65_{b}^{1} (0.11)	3.87_{b}^{1} (0.11)	3.21_{a}^{2} (0.10)	2.75_{a}^{1} (0.11)	2.67_{a}^{2} (0.11)	2.64_{a}^{2} (0.10)
High	4.23_{c}^{2} (0.10)	4.04_{b}^{1} (0.14)	2.53_{a}^{1} (0.10)	2.50_{a}^{1} (0.10)	2.25_{a}^{1} (0.14)	2.33_{a}^{1} (0.10)

Note. Means with different subscripts differ significantly from each other for a given dimension across condition. Means with different superscripts differ significantly from each other for a given dimension within condition.

(1, 246) = 43.80, $p < .001$, $\eta_p^2 = .15$. Participants agreed more with the non-discrimination norm in the non-discrimination norm ($M_{Study3a} = 5.26$, $SD = 1.70$; $M_{Study3b} = 5.09$, $SD = 1.32$) than in the honesty norm condition ($M_{Study3a} = 3.27$, $SD = 1.48$; $M_{Study3b} = 3.90$, $SD = 1.49$). In Study 3b, participants also clearly remembered the message presented in the survey they had received, $M = 5.93$, $SD = 1.20$ in the non-discrimination norm condition and $M = 2.86$, $SD = 1.96$ in the honesty norm condition, $F_{Study3b}(1, 246) = 223.05$, $p < .001$, $\eta_p^2 = .48$.

Warmth and Competence Judgments. Here, we only report the result pertaining to the interaction targeted by our hypothesis, that is, condition by dimension by external motivation to respond without prejudice for Studies 3a and 3b, and the condition by dimension by self-monitoring for Study 3b. The full results can be found in the Supplementary Materials.

Study 3a. We submitted the competence and warmth scores of the five groups to a mixed-model analysis using norm condition (with C1 opposing the control and the non-discrimination norm conditions, both = –1/3, to the honesty norm condition = 2/3; and C2 with non-discrimination condition = –1/2, control condition = 1/2, and the honesty norm condition = 0), dimension (warmth = –1/2 and competence = 1/2), (centered) external and (centered) internal motivation to respond without prejudice as our fixed effects and participants and target groups as our random effects. Importantly, our C1 norm condition contrast reflects our main prediction that the honesty norm condition should differ from the two other conditions.

The predicted three-way interaction involving condition, dimension, and external motivation was significant, $F(2, 1208) = 16.70$, $p < .001$, $\eta_p^2 = .03$. To explore the nature of this interaction, we relied on our two condition contrasts. The C2 by dimension interaction, the C2 by external motivation, as well as the three-way interaction involving C2 failed to reach significance, ps $> .10$. In sharp contrast, all these interactions proved significant when considering C1. Specifically, the significant C1 by dimension interaction, $F(1, 1208) = 112.52$, $p < .001$, $\eta_p^2 = .09$, confirmed that the decrease in the evaluation due to the honesty norm was more substantial for warmth ($M = 3.95$, $SE = 0.06$, and 2.87, $SE = 0.08$), $t(246) = –12.20$, $p < .001$, $\eta_p^2 = .38$, than for competence ($M = 2.54$, $SE = 0.06$, and 2.48, $SE = 0.08$), $t(246) = –0.66$, *ns.* (see Table 1). Next, the C1 by external motivation interaction, $F(1, 124) = 11.07$, $p < .002$, $\eta_p^2 = .08$, confirmed that the negative impact of honesty norm on judgments was less marked among low externally motivated ($M = 3.24$, $SE = 0.07$, and 2.93, $SE = 0.09$), $t(124) = –2.96$, $p < .004$, $\eta_p^2 = .07$, than among high externally motivated ($M = 3.25$, $SE = 0.08$, and 2.42, $SE = 0.09$), $t(124) = –7.56$, $p < .001$, $\eta_p^2 = .32$.

Finally, the critical C1 by dimension by external motivation interaction came out significant, $F(1, 1208) = 29.02$, $p < .0001$, $\eta_p^2 = .02$.

Decomposing this interaction by dimension and turning to warmth first, there was a significant C1 by external motivation interaction, $F(1, 246) = 32.74$, $p < .001$, $\eta_p^2 = .12$, confirming the presence of a greater difference between the non-discrimination norm and control conditions on the one hand and the honesty norm condition on the other among high externally motivated participants ($M = 4.13$, $SE = 0.09$, and 2.53, $SE = 0.10$), $t(246) = –12.33$, $p < .001$, $\eta_p^2 = .38$, than among low externally motivated ones ($M = 3.76$, $SE = 0.08$, and 3.21, $SE = 0.10$), $t(246) = –4.39$, $p < .001$, $\eta_p^2 = .07$. Conversely, the honesty norm led high externally motivated participants to evaluate the groups as colder than low externally motivated did, $t(246) = –5.02$, $p < .001$, $\eta_p^2 = .09$, whereas the two other conditions led to a more lenient judgment by high externally motivated than low externally motivated participants, $t(246) = 2.99$, $p < .001$, $\eta_p^2 = .03$. The analysis for the competence ratings only revealed a main effect of external motivation, $F(1, 246) = 12.20$, $p < .001$, $\eta_p^2 = .05$, in that low externally motivated individuals again saw the groups as more competent ($M = 2.69$, $SE = 0.07$) than high externally motivated did ($M = 2.36$, $SE = 0.07$).

Apart from an unpredicted significant C1 by external motivation by internal motivation, $F(1, 124) = 4.50$, $p < .04$, $\eta_p^2 = .04$, none of the other three-way interactions

Table 2. Means (Standard Errors) of Warmth and Competence as a Function of Condition and External Motivation to Respond Without Prejudice/Self-Monitoring (Study 3b).

	Dimension			
	Warmth		Competence	
Condition	Non-discr.	Honesty	Non-discr.	Honesty
External motivation				
Low	3.91_b^2 (0.16)	3.72_b^2 (0.17)	3.23_a^2 (0.16)	3.47_a^2 (0.17)
High	3.88_b^2 (0.16)	2.95_a^1 (0.16)	3.12_a^2 (0.16)	2.95_a^1 (0.16)
Self-monitoring				
Low	4.00_b^2 (0.16)	3.79_b^2 (0.16)	3.35_b^2 (0.16)	3.57_b^2 (0.16)
High	3.85_b^2 (0.17)	2.80_a^1 (0.16)	3.11_b^2 (0.17)	2.81_a^1 (0.16)

Note. Means with different subscripts differ significantly from each other for a given dimension across condition. Means with different superscripts differ significantly from each other for a given dimension within condition.

involving internal motivation to respond without prejudice reached significance, all $Fs < 1$, *ns*. Also, in line with what emerged in Study 2, the two four-way interactions (with C1 and C2) failed to reach significance, $F(1, 1208) = 2.29$, $p < .13$ and $F(1, 1208) = 1.79$, $p < .18$.

Study 3b

Internal and External Motivation to Respond Without Prejudice. We submitted the competence and warmth scores of the seven groups to a mixed-model analysis using norm condition (with non-discrimination = –1/2 and the honesty norm = 1/2), dimension (warmth = –1/2 and competence = 1/2), as well as external and internal motivation to respond without prejudice as our fixed effects and participants, groups, and traits as random effects.

As hypothesized, the condition by dimension by external motivation three-way interaction was significant, $F(1, 2699) = 7.08$, $p < .008$, $\eta_p^2 = .003$ (see Table 2). Decomposing this interaction by dimension, and as expected, the analysis for the warmth ratings revealed the presence of a significant condition effect, $F(1, 281) = 30.27$, $p < .001$, $\eta_p^2 = .10$, a significant effect of external motivation to respond without prejudice, $F(1, 284) = 13.90$, $p < .001$, $\eta_p^2 = .05$, as well as a significant condition by external motivation to respond without prejudice interaction, $F(1, 283) = 12.33$, $p < .001$, $\eta_p^2 = .04$. Simple effect analyses revealed that whereas no difference emerged between high and low externally motivated participants in the non-discrimination norm condition ($M = 3.88$, $SE = 0.16$, and $M = 3.91$, $SE = 0.16$), $t(282) = –0.18$, *ns*, the honesty norm condition led high externally motivated participants to rate the groups much colder than low externally motivated did ($M = 2.95$, $SE = 0.16$, and $M = 3.72$, $SE = 0.17$), $t(284) = –4.81$, $p < .001$, $\eta_p^2 = .08$. Conversely, low externally motivated participants rated the groups equally warm in the non-discrimination norm condition than in the honesty norm condition, $t(282) = –1.28$, *ns*. In sharp contrast, high externally motivated evaluated the groups as being much warmer in the non-discrimination norm condition than in the honesty norm one, $t(282) = –6.34$, $p < .001$, $\eta_p^2 = .12$.

Turning to competence, the analysis only revealed the presence of a significant main effect of external motivation to respond without prejudice, $F(1, 284) = 8.90$, $p < .003$, $\eta_p^2 = .03$, with low externally motivated participants giving higher competence ratings ($M = 3.35$, $SE = 0.14$) than high externally motivated did ($M = 3.04$, $SE = 0.14$).

Except for an unpredicted and barely significant dimension by external motivation by internal motivation, $F(1, 2699) = 3.86$, $p < .05$, $\eta_p^2 = .003$, none of the other three-way interactions involving internal motivation to respond without prejudice proved significant. Again, the four-way interaction was not significant, $F < 1$, *ns*.

Self-Monitoring. We conducted an exploratory analysis by submitting the competence and warmth scores about the seven groups to a mixed-model analysis using norm condition, dimension, self-monitoring as our fixed effects and participants, groups, and traits as random effects.

As we hoped, the hypothesized condition by dimension by self-monitoring three-way interaction proved significant, $F(1, 2703) = 7.67$, $p < .006$, $\eta_p^2 = .003$ (see Table 2). Again, decomposing this interaction by dimension, the analysis for the warmth ratings revealed a condition effect, $F(1, 279) = 37.07$, $p < .001$, $\eta_p^2 = .12$, an effect of self-monitoring, $F(1, 280) = 29.86$, $p < .001$, $\eta_p^2 = .10$, as well as a condition by self-monitoring interaction, $F(1, 281) = 15.70$, $p < .001$, $\eta_p^2 = .05$. Follow-up analyses revealed that no difference emerged between high and low self-monitors in the non-discrimination norm condition ($M = 3.85$, $SE = 0.17$, and $M = 4.00$, $SE = 0.16$), $t(281) = –0.98$, *ns*, whereas the honesty norm condition led high self-monitors to rate the groups much colder than low self-monitors did ($M = 2.80$, $SE = 0.16$, and $M = 3.79$, $SE = 0.16$), $t(280) = –7.13$, $p < .001$, $\eta_p^2 = .15$. Conversely, low self-monitors rated the groups equally warm in the non-discrimination and in the honesty norm condition, $t(280) = –1.45$, *ns*. In sharp

contrast, high self-monitors evaluated the groups much warmer in the non-discrimination than in the honesty norm condition, $t(280) = -7.06$, $p < .0001$, $\eta_p^2 = .15$.

The analysis on competence confirmed the presence of a main effect of self-monitoring, $F(1, 280) = 22.45$, $p < .001$, $\eta_p^2 = .07$. Low self-monitors rated the groups as more competent ($M = 3.46$, $SE = 0.14$) than high self-monitors did ($M = 2.96$, $SE = 0.15$). There was also a condition by self-monitoring interaction, $F(1, 281) = 5.90$, $p < .02$, $\eta_p^2 = .02$. Interestingly, when facing the non-discrimination norm, low self-monitors did not judge the groups more competent ($M = 3.35$, $SE = 0.16$) than high self-monitors did ($M = 3.11$, $SE = 0.17$), $t(281) = -1.52$, *ns*. In contrast, the honesty norm condition had high self-monitors rate the groups as less competent ($M = 2.81$, $SE = 0.16$) than low self-monitors did ($M = 3.57$, $SE = 0.16$), $t(280) = -5.42$, $p < .001$, $\eta_p^2 = .09$.

Discussion

The data from Studies 3a and 3b provided strong support for our hypotheses. In both studies and replicating the pattern observed in Study 2, participants evaluated the groups more harshly on competence than on warmth. As predicted, the confrontation with an honesty norm rather than a non-discrimination norm led participants to evaluate the groups less positively, especially on the warmth dimension. Crucially, this pattern emerged more strongly among high externally motivated participants and high self-monitor ones, than among their low externally motivated and low self-monitor counterparts.

General Discussion

Decades of research have shown that two fundamental dimensions, i.e., competence (the vertical dimension) and warmth (the horizontal dimension), orchestrate people's impressions about groups (Abele et al., 2021; Yzerbyt et al., 2025). While many groups in society have low status or limited resources, triggering judgments of low competence, some are viewed less negatively due to their perceived warmth, such as the elderly or people with disabilities (Cuddy et al., 2008; Fiske et al., 2002). In other words, these groups often face a paternalistic stereotype. The present research was built on the observation that such social groups are also seen as being socially protected. The question is thus whether these positive warmth judgments are genuine or simply reflect, at least for some people, an effort to conform to social norms. We hypothesized that while many social perceivers may truly endorse dominant social norms, this would be less the case for those who embrace paternalistic stereotypes for reasons of conformity rather than because of genuine internalization. In other words, warmth judgments about protected social groups may be less authentic among those who are motivated by external factors.

Four studies tested these ideas. Study 1 asked participants to evaluate the competence and warmth of groups located in different quadrants of the SCM model, and to assess the degree to which they thought the group was socially protected. As expected, the groups simultaneously rated low on competence and high on warmth were also socially protected.

Study 2 offered the first formal test of our hypothesis that some participants may express paternalistic views about protected groups not so much because they genuinely endorse these stereotypes but rather because they perceive them as socially enforced. We first assessed our participants' motivation to respond without prejudice. Then, we manipulated the nature of the normative pressure. Participants in the control condition completed a LDT for two protected groups. Following classic bogus pipeline instructions, participants in the bogus pipeline condition additionally learned that this specific task would reveal their true responses to the groups. All participants then conveyed their views about the groups using self-report questionnaires. Consistent with their placement in the SCM paternalistic quadrant, the groups generally came across as less competent than warm. More importantly, in line with our hypothesis, the bogus pipeline procedure only had an impact on the ratings of warmth and among high externally motivated individuals.

To increase the ecological validity of our test, we changed gears entirely in Studies 3a and 3b. Study 3a manipulated the degree to which participants would feel compelled to express their genuine views about five social groups by exposing participants to different norms, i.e., a non-discrimination norm versus an honesty norm versus a control condition (see Cambon & Yzerbyt, 2018). Fully replicating the pattern observed in Study 2, the honesty norm had a negative impact, especially on warmth judgments and for high externally motivated participants. The pattern in the control condition was the same as in the non-discrimination norm condition, confirming the ubiquity of the latter norm in real life.

Dropping the control condition and using all the available power to differentiate our two focal experimental conditions, Study 3b involved an even more diverse set of protected groups, a richer list of traits to measure competence and warmth, the vertical and horizontal dimensions, respectively, and a more general measure of sensitivity to the ambient social expectations, namely the self-monitoring scale (Snyder, 1974). The data again supported our predictions. The findings with the self-monitoring scale neatly echoed the results obtained with the IEM. This result is telling because the self-monitoring scale taps the extent to which people consider their attitudes as being socially acceptable in the specific context that they are facing. In other words, for high self-monitoring individuals, the reported attitudes very much reflect their perception of transient social requirements (Kardes et al., 1986; Snyder & Kendzierski, 1982). Such an inclination toward expressing socially guided attitudes proves more general than what the attitudes in favor of greater intergroup equality revealed by the IEM.

Overall, our data consistently showed that the paternalistic stereotypes associated with socially protected groups are not universally shared. Individuals who are externally motivated to express positive views about the warmth of certain groups do not necessarily hold these paternalistic stereotypes authentically. When they believe that their "real" views can be revealed or are expected to express their true beliefs, their judgments shift in a significant and predictable way.

Our findings may help explain why members of protected groups continue to face discrimination, despite expectations of a more positive stance from the majority about these groups. Indeed, it appears that some individuals may not genuinely endorse the positive stereotypes that often emerge when people complete trait questionnaires (Fiske et al., 2002) or report their so-called spontaneous impressions (Nicolas et al., 2021; Nicolas & Fiske, 2023). On a related note, these results also call into question the strength of anti-discrimination norms, which appear susceptible to being overcome by counter-discourse. Recent studies have demonstrated how campaign speeches and, more broadly, political agendas can undermine anti-discrimination standards and usher in a normative climate that favors the expression of prejudice (Ruisch & Ferguson, 2023). For example, Crandall et al. (2018) showed how the 2016 US political campaign increased the acceptability of prejudice toward groups, mainly paternalistic ones, targeted by the campaign. Clearly, the growing antagonism between the anti-discrimination norms and the counter-normative narrative of populists surging in many contemporary societies requires resolution. This is especially important given the consequences of expressing prejudice on the well-being and capability of marginalized, and mostly paternalistic, groups (Hu et al., 2024; Stanke et al., 2024). One avenue for future progress could be to highlight "all-inclusive" pro-diversity norms through interventions that target people's perception of social norms (Jansen et al., 2015; Murrar et al., 2020).

The present work also sheds light on the distinct nature of the horizontal dimension in social judgment. Several empirical efforts suggest that the horizontal dimension appears more "negotiable" than the vertical one (Cambon & Yzerbyt, 2018; Yzerbyt & Cambon, 2017) or, at least, that it is appraised more critically than the vertical dimension and often incorporates additional considerations (Koch et al., 2020). These two dimensions likely represent different realities. The vertical dimension reflects the societal status of groups, conveying information that primarily pertains to the targets and is less open to discussion. In contrast, the horizontal dimension involves factors that allow the perceiver's relationship with the target groups to influence the judgment. Therefore, consensus is likely to be lower on the horizontal dimension than on the vertical one. One symptom of this difference is that our manipulations, be it the bogus pipeline or the descriptive norms, had more of an impact on warmth than on competence judgments.

Our empirical efforts do have limitations. The design did not include a comparison with groups that are less targeted by social protection, which would have been useful in understanding whether the same discrepancy exists between the "normal" and "genuine" responses for groups that are high on the vertical dimension but low on the horizontal one. In general, people often describe bankers, rich people, senior executives, Asians, Jews, Germans, and the like negatively on horizontal traits, as confirmed in Study 1. However, it is unclear why some people might be hypocritical when they also assign high levels of vertical traits to these groups (Yzerbyt et al., 2022). After all, the very representation of these groups as assertive and agentic can evoke both respect and envy. This suggests that the limitation may be more apparent than real, and it is likely that only groups in the low vertical/high horizontal quadrant experience some degree of duplicity in their social image.

Another limitation is that our data were collected only within the French context and are not gender balanced. Although cross-cultural research on the SCM suggests that social groups tend to be in similar quadrants globally (Cuddy et al., 2009) and the list of protected groups in Study 1 closely mirrors that reported by Crandall & colleagues (2002), one should replicate these findings in different cultural settings. In addition to the groups themselves, the prevalence of the non-discrimination norm in society could also influence the results. However, some groups are likely to be socially protected across most historical and geographical contexts. If so, it is likely that some individuals will hold more negative views than they publicly declare. Thus, expanding this research to other cultures is a valuable direction for future research. Also, taking greater care for a more gender-balanced study would ensure greater confidence in the robustness of the findings.

The present efforts focused on the combined role of normative context and individual differences in the motivation to respond to the normative constraints and how this played out to orient judgments about protected groups. Admittedly, a question that comes to mind is whether, next to individual differences, more structural factors may shape people's tendency to produce strategic rather than genuine social evaluations, especially on the horizontal dimension. Since protected groups clearly enjoy a low status, one obvious candidate is perceivers' social status. In line with this reasoning, the dimensional compensation model (Schmitz & Yzerbyt, 2022; Yzerbyt, 2018) has argued that the authenticity of the stereotypes expressed specifically by high-status respondents who judge low-status groups is likely open to question (Cambon & Yzerbyt, 2018). Interestingly, Koch and colleagues (2020) suggest that high-status people are hardly expected to produce flattering views about low-status people anyway, presumably because of the dissimilarity between the groups. Why and when high-status people see low-status people as being warm thus stands as a fascinating question for future work.

Another issue concerns the fact that, in line with a long tradition of work on social evaluation, the present research essentially focused on the Big Two dimensions. Next to the SCM that occupies a unique position in the field, several other models build upon the same or a closely related

distinction (Abele et al., 2021; Koch et al., 2020). Rather than relying solely on the Big Two dimensions, more recent efforts now tend to make distinctions between several facets (for a recent review, see Yzerbyt et al., 2025). Next to a vertical dimension that encompasses assertiveness and ability as two facets of competence, the horizontal dimension comprises two friendliness and morality as two different facets of warmth (Brambilla et al., 2012, 2021). One promising development of the present efforts would be to build upon these nuances and examine which one of these two facets of warmth is more likely to be associated with less genuine responses on the part of social perceivers. This is certainly one item that stands very high on our research agenda.

To sum up, while the social protection of certain groups can be seen as positive, there are drawbacks to being placed in the low vertical/high horizontal quadrant, and such representations can reinforce the status quo. Even aside from the negative impact of these paternalistic stereotypes, the positive traits attributed to these groups may not be universally accepted, and membership in these groups does not guarantee freedom from prejudice and discrimination. One reason for this may be that the norm of non-discrimination, which ideally shapes people's reactions, is not fully embraced by those who merely conform for external reasons. Our research shows that positive views about the warmth of socially protected groups may not always be sincere. As a result, when people claim that elderly individuals, the poor, or people with disabilities are kind and deserving of compassion, only some may truly act on these beliefs.

ORCID iDs

Vincent Yzerbyt https://orcid.org/0000-0003-1185-4733
Odile Rohmer https://orcid.org/0000-0003-3998-7570
Laurent Cambon https://orcid.org/0000-0003-3493-3944

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The present research has been partially supported by a Belgian National Fund for Scientific Research FNRS grant PDR T.0253.19 and by a Templeton World Charities Foundation Grant TWCF-2022-30553 to the first author.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Supplemental material

Supplemental material is available online with this article.

Notes

1. As part of another project, participants were also asked to fill in a French version of the Social Dominance Orientation at the end of the session. We also measured a series of additional aspects linked to the groups, such as status and perceived realistic and symbolic threat (Kervyn et al., 2015). The values associated with each group are presented in the Supplementary Materials.
2. We initially only considered an additive model. For this reason, the selection of our groups in Studies 2, 3a, and 3b rested on the difference between competence and warmth. We thank one reviewer for suggesting the inclusion of the interaction term in the model as this further improved its explanatory power.
3. The same factorial structure was obtained in all subsequent experiments (variance explained by the two factors ranging from 59% to 80%, r between the factors ranging from –.14 up to –.44, $\alpha_{external}$ ranging from .85 to .94, $\alpha_{internal}$ ranging from .74 to .92. Therefore, we do not report these data anymore (see Supplementary materials). In Study 3a, one item tapping the internal motivation was lost due to a program malfunction, with no impact on the reliability of the scale.
4. The PANGEA webapp (short for Power ANalysis for GEneral ANOVA designs) is an open-source, user-friendly online tool designed for conducting power analyses in complex ANOVA designs commonly used in psychology and neuroscience. PANGEA fills a major gap by offering a robust, accessible solution for power analysis in multifactor experimental designs beyond the capabilities of standard tools like G*Power (which often only handle simpler analyses).
5. As such, our design included two continuous factors, namely external and external motivation to respond without prejudice. Because PANGEA treats all factors as categorical, the power of our design is underestimated. The same consideration applies to all subsequent studies.
6. As a part of another project, participants were also asked to fill in six scales measuring negative emotions toward the self at the end of the experimental session.
7. Because we initially treated our two groups as replications, we did not include the target group as a factor in our power analysis. However, we also considered using it as a fixed factor in our analysis to keep the distinction between elderly people and people with disabilities. Doing this revealed no significant effect, except for an unpredicted interaction with external motivation to respond without prejudice ($F = 4.28$, $p < .04$). To simplify the presentation of our results, we dropped this factor from the analyses.
8. We initially conducted our power calculations and our analyses by treating the target group as a fixed factor. Apart from a main effect ($F = 7.78$, $p < .012$), no other effects involving the target group factor proved significant. Rather than dropping this factor altogether, we decided to treat both the target group and the participants as random effects. Although doing so decreases power substantially (.30), it also ensures generalization from our findings.

References

Abele, A. E., Ellemers, N., Fiske, S. T., Koch, A., & Yzerbyt, V. (2021). Navigating the social world: Toward an integrated framework for evaluating self, individuals, and groups. *Psychological Review*, *128*(2), 290–314. https://doi.org/10.1037/rev0000262

Abele, A. E., & Wojciszke, B. (2014). Communal and agentic content in social cognition: A dual perspective model. In *Advances in Experimental Social Psychology* (Vol. 50, pp. 195–255). Academic Press.

Bergsieker, H. B., Leslie, L. M., Constantine, V. S., & Fiske, S. T. (2012). Stereotyping by omission: Eliminate the negative, accentuate the positive. *Journal of Personality and Social Psychology*, *102*(6), 1214–1238. https://doi.org/10.1037/a0027717

Brambilla, M., Sacchi, S., Rusconi, P., Cherubini, P., & Yzerbyt, V. Y. (2012). You want to give a good impression? Be honest! Moral traits dominate group impression formation. *British Journal of Social Psychology*, *51*, 149–166. https://doi.org/10.1111/j.2044-8309.2010.02011.x

Brambilla, M., Sacchi, S., Rusconi, P., & Goodwin, G. (2021). The primacy of morality in impression development: Theory, research, and future directions. *Advances in Experimental Social Psychology*, *64*, 187–262. https://doi.org/10.1016/bs.aesp.2021.03.001

Cambon, L., & Yzerbyt, V. Y. (2018). Two routes toward compensation: An investigation into the mechanisms of compensation for high-and low status groups. *Journal of Experimental Social Psychology*, *77*, 24–35. https://doi.org/10.1016/j.jesp.2018.04.002

Cialdini, R. B., Kallgren, C. A., & Reno, R. R. (1991). A focus theory of normative conduct: A theoretical refinement and reevaluation of the role of norms in human behavior. *Advances in Experimental Social Psychology*, *24*, 201–234. https://doi.org/10.1016/S0065-2601(08)60330-5

Clément-Guillotin, C., Rohmer, O., Forestier, C., Guillotin, P., Deshayes, M., & d'Arripe-Longueville, F. (2018). Implicit and explicit stereotypes content associated to persons with disability: Does sport can change anything? *Psychology of Sport and Exercise*, *38*, 192–201. https://doi.org/10.1016/j.psychsport.2018.06.014

Crandall, C. S., Eshleman, A., & O'Brien, L. (2002). Social norms and the expression and suppression of prejudice: The struggle for internalization. *Journal of Personality and Social Psychology*, *82*(3), 359–378. https://doi.org/10.1037/0022-3514.82.3.359

Crandall, C. S., Miller, J. M., & White, M. H. (2018). Changing norms following the 2016 U.S. presidential election: The Trump effect on prejudice. *Social Psychological and Personality Science*, *9*(2), 186–192. https://doi.org/10.1177/1948550617750735

Cuddy, A. J., Fiske, S. T., & Glick, P. (2008). Warmth and competence as universal dimensions of social perception: The stereotype content model and the BIAS map. *Advances in Experimental Social Psychology*, *40*, 61–149. https://doi.org/10.1016/S0065-2601(07)00002-0

Cuddy, A. J., Fiske, S. T., Kwan, V. S., Glick, P., Demoulin, S., Leyens, J. P., & . . .Ziegler, R. (2009). Stereotype content model across cultures: Towards universal similarities and some differences. *British Journal of Social Psychology*, *48*(1), 1–33. https://10.1348/014466608X314935

Czopp, A. M., Kay, A. C., & Cheryan, S. (2015). Positive stereotypes are pervasive and powerful. *Perspectives on Psychological Science*, *10*(4), 451–463. https://doi.org/10.1177/1745691615588091

Dunton, B. C., & Fazio, R. H. (1997). An individual difference measure of motivation to respond without prejudiced reactions. *Personality and Social Psychology Bulletin*, *23*(3), 316–326. https://doi.org/10.1177/0146167297233009

Fiske, S. T., Cuddy, A. J., & Glick, P. (2007). Universal dimensions of social cognition: Warmth and competence. *Trends in Cognitive Sciences*, *11*(2), 77–83. https://doi.org/10.1016/j.tics.2006.11.005

Fiske, S. T., Cuddy, A. J., 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

Gana, G., & Brechenmacher, N. (2001). Structure latente et validité de la version française du self-monitoring scale: Échelle de monitorage de soi [Latent structure and validity of the French version of the self-monitoring scale]. *Année Psychologique*, *101*(3), 393–420. https://doi.org/10.3406/psy.2001.1338

Granjon, M., Rohmer, O., Popa-Roch, M., Aubé, B., & Sanrey, C. (2023). Disability stereotyping is shaped by stigma characteristics. *Group Processes & Intergroup Relations*, *27*(6), 1403–1422. https://doi.org/10.1177/13684302231208534

Hu, Y., Purol, S. M., Shen, Y., & Zheng, Y. (2024). Perceived racism and well-being in university racial/ethnic minority students: The complex roles of racial/ethnic identity and self-esteem. *Current Psychology*, *43*, 37196–37207. https://doi.org/10.1007/s12144-024-07099-7

Ilmarinen, V. J., Isotalo, V., Lönnqvist, J. E., & von Schoultz, Å. (2022). Do politicians' answers to voting advice applications reflect their sincere beliefs? Comparing publicly and confidentially stated ideological positions in a candidate-centred electoral context. *Electoral Studies*, *79*, 102504.

Jansen, W. S., Otten, S., & van der Zee, K. I. (2015). Being part of diversity: The effects of an all-inclusive multicultural diversity approach on majority members' perceived inclusion and support for organizational diversity efforts. *Group Processes & Intergroup Relations*, *18*(6), 817–832. https://doi.org/10.1177/1368430214566892

Jeffries, C. H., Hornsey, M. J., Sutton, R. M., Douglas, K. M., & Bain, P. G. (2012). The David and Goliath principle: Cultural, ideological, and attitudinal underpinnings of the normative protection of low-status groups from criticism. *Personality and Social Psychology Bulletin*, *38*, 1053–1065. https://doi.org/10.1177/0146167212444454

Jones, E. E., & Sigall, H. (1971). The bogus pipeline: A new paradigm for measuring affect and attitude. *Psychological Bulletin*, *76*(5), 349–364. https://doi.org/10.1037/h0031617

Kardes, F. R., Sanbonmatsu, D. M., Voss, R. T., & Fazio, R. H. (1986). Self-monitoring and attitude accessibility. *Personality and Social Psychology Bulletin*, *12*(4), 468–474. https://doi.org/10.1177/0146167286124010

Kay, A. C., Day, M. V., Zanna, M. P., & Nussbaum, A. D. (2013). The insidious (and ironic) effects of positive stereotypes. *Journal of Experimental Social Psychology*, *49*(2), 287–291. http://dx.doi.org/10.1016/j.jesp.2012.11.003

Kervyn, N., Fiske, S. T., & Yzerbyt, V. Y. (2015). Forecasting the primary dimension of social perception: Symbolic and realistic threats together predict warmth in the stereotype content model. *Social Psychology*, *46*, 36–45. https://doi.org/10.1027/1864-9335/a000219

Klein, O., Snyder, M., & Livingston, R. W. (2004). Prejudice on the stage: Self-monitoring and the public expression of group attitudes. *British Journal of Social Psychology*, *43*(2), 299–314. https://doi.org/10.1348/0144666041501697

Koch, A., Imhoff, R., Unkelbach, C., Nicolas, G., Fiske, S., Terache, J., Carrier, A., & Yzerbyt, V. Y. (2020). Groups' warmth is a personal matter: Understanding consensus on stereotype dimensions reconciles adversarial models of social evaluation. *Journal of Experimental Social Psychology*, *89*, Article 103995. https://doi.org/10.1016/j.jesp.2020.103995

Murrar, S., Campbell, M. R., & Brauer, M. (2020). Exposure to peers' pro-diversity attitudes increases inclusion and reduces the achievement gap. *Nature Human Behaviour*, *4*(9), 889–897. https://doi.org/10.1038/s41562-020-0899-5

Nicolas, G., Bai, X., & Fiske, S. T. (2021). Comprehensive stereotype content dictionaries Using a semi-automated method. *European Journal of Social Psychology*, *51*(1), 178–196. https://doi.org/10.1002/ejsp.2724

Nicolas, G., & Fiske, S. T. (2023). Valence biases and emergence in the stereotype content of intersecting social categories. *Journal of Experimental Psychology: General*, *152*(9), 2520–2543. https://doi.org/10.1037/xge0001416

Nier, J. A. (2005). How dissociated are implicit and explicit racial attitudes? A bogus pipeline approach. *Group Processes & Intergroup Relations*, *8*(1), 39–52. https://doi.org/10.1177/1368430205048615

Plant, E. A., & Devine, P. G. (1998). Internal and external motivation to respond without prejudice. *Journal of Personality and Social Psychology*, *75*(3), 811–832. https://doi.org/10.1037/0022-3514.75.3.811

Plant, E. A., & Devine, P. G. (2001). Responses to other-imposed pro-Black pressure: Acceptance or backlash? *Journal of Experimental Social Psychology*, *37*(6), 486–501. https://doi.org/10.1006/jesp.2001.1478

Roese, N. J., & Jamieson, D. W. (1993). Twenty years of bogus pipeline research: A critical review and meta-analysis. *Psychological Bulletin*, *114*(2), 363–375. https://doi.org/10.1037/0033-2909.114.2.363

Rohmer, O., & Louvet, E. (2012). Implicit and explicit measures of the stereotype content associated with disability. *British Journal of Social Psychology*, *51*(4), 732–740. https://doi.org/10.1177/1368430216638536

Rohmer, O., & Louvet, E. (2018). Implicit stereotyping against people with disability. *Group Processes & Intergroup Relations*, *21*(1), 127–140. https://org/10.1177/1368430216638536

Rohmer, O., Palomares, E. A., & Popa-Roch, M. (2024). Attitudes towards disability and burnout among teachers involved in inclusive education. *International Journal of Disability, Development and Education*, *71*(1), 83–100. https://doi.org/10.1080/1034912X.2022.2092078

Ruisch, B. C., & Ferguson, M. J. (2023). Did Donald Trump's presidency reshape Americans' prejudices? *Trends in Cognitive Sciences*, *27*(3), 207–209. https://doi.org/10.1016/j.tics.2022.12.013

Schmitz, M., & Yzerbyt, V. (2022). Direct and indirect dimensional compensation: Is there a difference between observers and group members? *Group Processes & Intergroup Relations*, *25*(2), 335–353. https://doi.org/10.1177/1368430220963176

Snyder, M. (1974). Self-monitoring of expressive behavior. *Journal of Personality and Social Psychology*, *30*(4), 526–537. https://doi.org/10.1037/h0037039

Snyder, M., & Kendzierski, D. (1982). Acting on one's attitudes: Procedures for linking attitude and behavior. *Journal of Experimental Social Psychology*, *18*(2), 165–183. https://doi.org/10.1016/0022-1031(82)90048-8

Stangor, C., Sechrist, G. B., & Jost, J. T. (2001). Changing racial beliefs by providing consensus information. *Personality and Social Psychology Bulletin*, *27*(4), 486–496. https://doi.org/10.1177/0146167201274009

Stanke, F. A., Kuper, N., Fetz, K., & Echterhoff, G. (2024). Discriminatory, yet socially accepted? Targets' perceptions of subtle and blatant expressions of ethno-racial prejudice. *Frontiers in Social Psychology*, *2*, Article 1343514. https://doi.org/10.3389/frsps.2024.1343514

Tourangeau, R., Smith, T. W., & Rasinski, K. A. (1997). Motivation to report sensitive behaviors on surveys: Evidence from a bogus pipeline experiment. *Journal of Applied Social Psychology*, *27*(3), 209–222. https://doi.org/10.1111/j.1559-1816.1997.tb00629.x

Yzerbyt, V. Y. (2018). The dimensional compensation model: Reality and strategic constraints on warmth and competence in intergroup perceptions. In A. E. Abele & B. Wojciszke (Eds.), *Agency and communion in social psychology* (pp. 126–141). Routledge. https://doi.org/10.4324/9780203703663

Yzerbyt, V. Y., Barbedor, J., Carrier, A., & Rohmer, O. (2022). The facets of social hierarchy: How judges' legitimacy beliefs and relative status shape their evaluation of assertiveness and ability. *International Review of Social Psychology*, *35*(1). https://doi.org/10.5334/irsp.695

Yzerbyt, V. Y., & Cambon, L. (2017). The dynamics of compensation: When ingroup favoritism paves the way for outgroup praise. *Personality and Social Psychology Bulletin*, *43*, 587–600. https://doi.org/10.1177/0146167216689066

Yzerbyt, V. Y., Koch, A., Brambilla, M., Ellemers, N., Fiske, S. T., & Nicolas, G. (2025). Dimensions of stereotypes about groups. *Current Directions in Psychological Science*. Advance online publication. https://doi.org/10.1177/09637214251345489