INTRODUCTION

The number of women who finish an academic degree is rising, with female students receiving higher numbers of academic degrees than male students (Eurostat, 2018). Hence, the number of women professionals is also rising (Eurostat, 2018). Nevertheless, women are still underrepresented in leadership positions (European Commission, 2017). Similarly, women are rarely found in the position of professors (European Commission, 2015). This finding has been consistent for years (e.g., Finkelstein et al., 2016). Thus, there is much research investigating the underlying causes (for an overview in psychology see: Gruber et al., 2020). However, there is no conclusive agreement on the causes of these differences. While some studies indicate a role of the self-selection of the career path (Ceci & Williams, 2011; Clark et al., 2016); others indicate that discrimination in hiring decisions may also play a role (Moss-Racusin et al., 2012). One recurring reason proposed in the literature is that gender stereotypes facilitate discriminatory decisions even in the absence of differences in work commitment or performance (Berkery et al., 2013; Duehr & Bono, 2006; Heilman et al., 1989; Koenig et al., 2011; Schein, 1994).
We define gender stereotypes as socially shared generalized beliefs about the attributes of women and men as a group (Sczesny et al., 2018). These attributes are often described along two dimensions: agency and communion. Agency refers to an individual’s motivation for self-assertion, environmental mastery, competence, and independence. It contains traits like potent, dominant, competitive, analytical, independent, and influential. Communion refers to a motivation to search for relations with others and comprehends traits like warm, friendly, helpful, and well-intentioned (Bakan, 1966; Carli et al., 2016; Imhoff & Koch, 2017).

### 1.1 | The think manager—think male effect

To further understand disparities in one area, it may be helpful to look at other areas where it is understood better. Agentic attributes are more strongly associated with men than with women, whereas communal attributes are more strongly associated with women than with men (Bakan, 1966; Eagly, 2013; Eagly & Johannesen-Schmidt, 2001; Eagly & Karau, 2002; Eagly et al., 2020). Likewise, managers are usually more associated with agentic attributes and therefore, are more strongly connected to attributes associated with men rather than women. This overlap between the attributes associated with men and managers has first been examined by Schein (1973, 1975). Thus, both male and female managers perceive typical managers as more similar to men than to women. This result is called the “think manager—think male” effect (Schein & Mueller, 1992, see also Sczesny, 2003). The think manager—think male effect is often proposed as a reason for the apparent discrimination of women when they choose the career path of a manager.

The think manager—think male and has been frequently replicated (Berkery et al., 2013; Duehr & Bono, 2006; Heilman et al., 1989; Koenig et al., 2011). Furthermore, this effect is mostly culture-independent. It was also found (at least for male participants) in samples from Great Britain and Germany (Schein & Mueller, 1992), Japan and China (Schein et al., 1996), Egypt (Moneim Elsaid & Elsaid, 2012), Saudi Arabia (Marmenout, 2009), Turkey (Fullagar et al., 2003; Gupta et al., 2005), Sweden (Fullagar et al., 2003), India (Gupta et al., 2005), Canada (Norris & Wylie, 1995; Orser, 1994), New Zealand (Sauers et al., 2002), Hawai’i (de Pillis et al., 2008), and South Africa (Booysen & Nkomo, 2010). Although the number of women working as leaders has increased (e.g., Adler, 2002), the think manager—think male effect is robust, at least for male raters. For female raters, however, the think manager—think male effect has been found to decrease (e.g., Berkery et al., 2013). Yet, male participants still rated managers more similarly to men than to women. Female participants, however, rated managers similarly to both men and women (Berkery et al. 2013; Duehr & Bono, 2006; Koenig et al., 2011).

Since women are also underrepresented in (senior) scientific positions, one may also assume the discrimination of women when applying for these positions (e.g., Moss-Racusin et al., 2012). In general, men are stronger associated with science than women (for a meta-analysis see: Miller et al., 2018). Scientists also are, like men, rated highly on agentic attributes. They are also rated lower than either gender on communal attributes, making them even more unlike women, who are rated highly on communal attributes (Carli et al., 2016; Ramsey, 2017). This indicates that a think scientist—think male effect may be at play and might be based on similar attributes as the think manager—think male effect, which we consider in this research.

### 1.2 | How stereotypes of women and scientist are changing

The association between men and science has weakened over the years. Gruber et al. (2020) attribute this reduction to changes in the stereotypes about women. However, Eagly et al. (2020) found a change in the women stereotype to more communal attributes. Thus, if the discrimination of women in science is based on a discrepancy between the stereotype of scientists as un-communal and women as communal, one would expect associations between women and scientists to have decreased in the past. As the opposite is the case, that is associations between women and scientists have increased in recent years (Gruber et al., 2020; Miller et al., 2018), this indicates that the theorized think scientist—think male effect might work differently than to the think manager—think male effect. Miller et al. (2015) show that the increased associations between women and scientists likely stem from increased exposure to female scientists. Additionally, the increase of the association between women and scientists is much more pronounced for women and girls than for men or boys (Charlesworth & Banaji, 2019; Miller et al., 2018).

The increased exposure to scientists as a key factor may indicate that it is not so much the changes in the image of women or their self-image that has led to increased associations, but that the image of scientists may also play a role. Besides, since the increase is mainly driven by women or girls, it is important to also investigate which role the self-image of men and women plays in discrimination.

### 1.3 | Mechanisms of discrimination

The match or mismatch between attributes ascribed to men or women and a position may result in discrimination in multiple ways:

First, people may believe that women lack attributes they assume to be required for a position if these attributes are part of the male but not the female stereotype (descriptive stereotypes about women; Eagly & Karau, 2002). Second, there may be a mismatch between norms applied to women (prescriptive stereotypes), that is beliefs about the way women should behave, and the behavior that is expected from a person in a certain position (Eagly & Karau, 2002). These two types of discrimination may both also create some form of self-selection, through which this discrimination is internalized. Social cognitive career theory (SCCT; e.g., Lent & Brown, 2016) indicates that people will take a career path, through which they believe they have the best chances to attain their goals. Hence, persons
may choose not to pursue a certain career path, either because they do not see how this path aligns with their own goals (mismatch between goals and outcome expectations), or because they believe themselves to be incapable to reach the goals through this path (insufficient self-efficacy). Such internalized stereotypes have been shown to play a role in self-selection for scientific careers (Diekman et al., 2010; Tellhed et al., 2017). The third mechanism of discrimination arises if the attributes associated with women match a currently held position better (e.g., a subordinate position) than the desired one (e.g., a leadership position), thus leading others to believe the women to be a good fit for the current but not the desired position (sticky-floor effect; e.g., Braun et al., 2017). Again this type of discrimination may work both through external discrimination as well as through internalized self-selection, that is if women see themselves fitting the current position better than another one (see also Stoet & Geary, 2018; Wang et al., 2013, who showed that women tend to be “lured” away from STEM careers if they also show strong abilities required for non-STEM related fields). Because of these various ways in which stereotypes may lead to discrimination, it is not only important to identify which groups are considered similar or dissimilar but also to understand the detailed structure of the association, that is which attributes are considered similar and which are considered dissimilar between the groups. We investigate how the perceived attribute structure of women, men, managers, and scientists match.

Different types of discrimination require different interventions to reduce the obstacles for women. Most importantly, discrimination may be created due to the stereotype about the person applying for a job (person stereotype) or due to the stereotype about the position that is desired (role stereotype). For example, women believed themselves to be worse in math than men (Nosek et al., 2002), and this belief is internalized even when the actual mathematical abilities are comparable (Bench et al., 2015). Thus, women may be discriminated against externally when applying for positions requiring mathematical skills. Similarly, women could be strayed away from such a position due to their internalized belief. Since women’s and men’s actual mathematical abilities are comparable (Lindberg et al., 2010) this discrimination is driven by an incorrect person stereotype. An intervention against this type of discrimination must correct this person stereotype, either by boosting women’s confidence in their mathematical abilities or by highlighting the fact that the differences in mathematical abilities between genders are negligible. While discrimination due to descriptive stereotypes is driven through person stereotypes, discrimination due to prescriptive stereotypes or self-selection due to goal mismatches is driven by role stereotypes. Prescriptive stereotypes about women include the norm that women should be more communal than other groups (e.g., Eagly & Karau, 2002; Fiske et al., 2002). Thus, women may be hindered from taking positions, which do not fit a communal profile. Since science is often seen as not serving a communal goal (Clark et al., 2016), we assume that the resulting perceived lack-of-fit (Heilman et al., 1989) between women and scientists may hinder women from becoming successful scientists both due to external discrimination, if they are seen as not following their assigned roles, and through self-selection, if they feel their goals are unattainable through a career in science. This effect is not just driven by norms about women, but rather, the role stereotype of a scientist as unsocial also plays a key role (Clark et al., 2016). Changes in the role stereotype could, therefore, reduce discrimination.

1.4 Differences and similarities between managers and scientists

Unlike managers, who fulfill a clear leadership role, scientists must fulfill a wide range of different roles, depending on their current task, thus the role stereotype for scientists is less clear. While successful scientists are also leaders in the sense that they have to oversee a group of subordinate researchers, apply for grants, and set a research direction for their group, they are often also employed in teaching positions, are responsible for the advancement of their subordinate researchers (e.g., as Ph.D. students), have to find partners for research collaborations, and have to present their findings at conferences or in articles. Thus, while there may be a mismatch between some parts of the scientific work and the prescriptive stereotypes or goals of women, other parts may match. This leads to possible interventions that can work through changes in the role stereotype for scientists, instead of trying to change the person stereotype for women. Focusing on the teaching role of scientists can help women to better identify with scientists (Ramsey, 2020). Similarly, science is often seen to lack communal value and to serve mostly agentic goals, which can lead to women self-selecting themselves out of scientific careers due to the perceived mismatch (Diekman et al., 2010). Highlighting communal parts of the work of scientists can then increase the belief that communal goals of women may be achieved through a career in science (Clark et al., 2016; Ramsay, 2020), showing that stereotypes about scientists may be open to such a change. Also, some scientific tasks may be perceived gender neutral, for example, writing papers. However, it should be pointed out, that although these gender-neutral tasks may be seen as not hindering women directly, they also do not lead to communal goals, that is they only provide a poor possibility to work directly with other people. Clark et al. (2016) specifically employed images of scientists working alone versus scientists collaborating strongly with others and showed that the image of the collaborative researcher increased the perception of science as a communal endeavor.

A similar discussion can be found in the literature about the think manager—think male effect. Researchers discuss if changes in this effect (i.e., women rating both women and men similarly to managers in recent years) occur because person stereotypes of women have changed (Diekman & Eagly, 2000; Twenge, 1997) or because leader roles have changed (Koenig et al., 2011). In a meta-analysis, Haines et al. (2016) found that, while gender stereotypes are rather stable over time, the roles associated with women have become more diverse. This hints at discrimination via descriptive stereotypes in previous years, which has now been reduced because the stereotypes about the discriminated group have been
reduced. For Australian employees from 25 different companies, Griffiths et al. (2019) found that women were associated with communal and agentic attributes—whereas men were still only associated with mostly agentic traits—and that leadership was associated with both communal and agentic attributes, resulting in a better fit between women and leadership roles than men and leadership roles. This, however, indicates that discrimination via prescriptive stereotypes was responsible for the think manager—think male effect, and the reduction of this effect is caused by a reduced conflict between norms about women and the role of managers. This shows that it is not only necessary to discuss which groups are matching or not, but rather to have a detailed look at the mechanism through which the think manager—think male effect and the think scientist—think male effect lead to discrimination.

1.5 | The current study

We investigate which attributes people associate with successful managers as well as scientists. Specifically, we conceptually replicate the think manager—think male effect and the think scientist—think male effect, and test if these effects are both driven by the same attributes being associated with leadership. Thus, a broad match between the attributes associated with managers and scientists could imply that similar interventions can be used to counter discrimination against women in science and management. Differences between the attributes associated with managers and scientists, however, would indicate that different interventions are necessary to counteract either effect. Thus, a detailed analysis of the structure of the stereotype can provide insight into the mechanisms causing discrimination in the case of women as managers and women as scientists.

Based on the existing literature, we formed several hypotheses. We followed a Popperian approach, while forming our hypotheses (e.g., Glöckner & Betsch, 2011), thus we only formed hypotheses where we found these to correspond to justified and testable predictions based on existing theories. Since the think manager—think male effect was shown across many different settings, we expected to be able to replicate this effect here as well for German participants. Thus, we expected a significantly higher correlation between the ratings for men and managers on the Schein Descriptive Index (DI) than between women and managers (Hypothesis 1a). Furthermore, because scientists are also often male stereotyped, we expected a similar think scientist—think male effect, that should likewise appear as a higher correlation between men and scientists on the DI than between women and scientists (Hypothesis 2a). We expect this effect to be both observable when the overall correlations are investigated, as well as when the number of items that differ significantly between the groups is inspected. Thus, we expect women to differ significantly from managers on more items than men (Hypothesis 1b) and likewise, we expected women to differ significantly from scientists on more items than men (Hypothesis 2b).

While we formed the hypotheses as predictions based on existing theories in psychology, we also agree with Oberauer and Lewandowsky (2019) that psychology is in many cases still lacking sufficient theory to make detailed and testable predictions. Therefore, in addition to these hypotheses, we will provide exploratory analyses to provide insight into the structure of the stereotypes about men, women, managers, and scientists to explore whether or not the think manager—think male effect and the think scientist—think male effects are based on similar attributes. As these analyses are exploratory and not based on the same Popperian rigor we used when forming our predictive hypotheses, we explicitly abstained from presenting hypotheses about this exploratory part.

2 | METHOD

2.1 | Participants

A sample of 319 adults took part in the study. Like Berkery et al. (2013) we removed participants who showed signs of insufficient effort in responding (IER; Dunn et al., 2018). To detect IER we used the individual response variability index (IRV; Dunn et al., 2018) calculated using the careless R-package (Yentes & Wilhelm, 2018). Participants whose IRV was lower than 0.35 were considered insufficient effort responders. This value was chosen because it allowed us to remove approximately 10% of the participants (Dunn et al., 2018). The criterion led to the removal of the data from 28 participants (8.8%). Of the remaining 291 participants, 73.5% were female and 26.5% male with no significant differences in the number of male and female participants across the four conditions ($F(3;287) = 0.66; p = .58$ in a logistic regression). Their mean age was 30.7 years ($\text{Median} = 28; \text{SD} = 8.7; \text{Min} = 18; \text{Max} = 60$) with no significant differences in the ages across the four conditions ($F(3;287) = 2.28; p = .08$ in a linear regression).

One participant entered the age as 12 years, but as the instructions clearly stated that this study was only targeted at adults, we considered this a miss-entry (this was supported by the fact, that the same participant reported having finished secondary school and being enrolled as a student). We classified the free-text occupations according to ISCO-08 (International Labour Organization [ILO], 2012). The majority (121 participants) were not classifiable, because they were students, unemployed, or similar. Of the remaining participants, the largest subgroups were from healthcare (18 participants) or teaching-related positions, including teaching at a university (18 participants). A complete overview of all occupations by condition is provided in Table 1. 17 (5.8%) participants worked in management positions (no significant differences between conditions: $F(3;287) = 1.13; p = .34$ in a logistic regression) with an average of 7.9 years of work management experience ($\text{Median} = 4; \text{SD} = 8.4$) and 12 (4.1%) participants worked in scientific fields (no difference between conditions: $F(3;287) = 1.09; p = .35$ in a logistic regression) with an average of 5.1 years’
experience in a scientific position (Median = 4.5; SD = 3.5). A detailed overview of the demographic structure of our sample across the four conditions is provided in Table 2. Participants were recruited via the course credit system of the FernUniversität in Hagen. The students of this distance teaching university are highly diverse concerning age, political attitudes, family status, and occupation (Stürmer et al., 2018). University of Hagen students live all across Germany. About 80% are working or self-employed during their studies. Further participants were recruited via social media posts, targeting general population participants.

2.2 | Measurement instrument

We used the Schein Descriptive Index (DI) to measure the associated traits of women, men, successful scientists, and successful middle

| ISCO-08                                      | Condition | Men | Women | Managers | Scientists | Total |
|----------------------------------------------|-----------|-----|-------|----------|------------|-------|
| OC02 (Non-commissioned armed forces officers) | 0         | 0   | 0     | 1        | 1          |       |
| OC11 (Chief executives, senior officials and legislators) | 0         | 0   | 0     | 1        | 1          |       |
| OC12 (Administrative and commercial managers) | 0         | 3   | 2     | 2        | 7          |       |
| OC22 (Health professionals)                  | 3         | 4   | 8     | 3        | 18         |       |
| OC23 (Teaching professionals)                 | 3         | 3   | 6     | 6        | 18         |       |
| OC24 (Business and administration professionals) | 1         | 3   | 0     | 0        | 4          |       |
| OC25 (Information and communications technology professionals) | 1         | 1   | 1     | 1        | 4          |       |
| OC26 (Legal, social and cultural professionals) | 3         | 2   | 2     | 5        | 12         |       |
| OC32 (Health associate professionals)         | 1         | 0   | 0     | 0        | 1          |       |
| OC33 (Business and administration associate professionals) | 1         | 6   | 5     | 3        | 15         |       |
| OC34 (Legal, social, cultural and related associate professionals) | 4         | 2   | 3     | 3        | 12         |       |
| OC41 (General and keyboard clerks)            | 0         | 1   | 1     | 1        | 3          |       |
| OC42 (Customer services clerks)               | 0         | 1   | 2     | 0        | 3          |       |
| OC43 (Numerical and material recording clerks) | 0         | 1   | 1     | 0        | 1          |       |
| OC51 (Personal service workers)               | 1         | 1   | 1     | 2        | 5          |       |
| OC52 (Sales workers)                          | 1         | 0   | 0     | 1        | 2          |       |
| OC54 (Protective services workers)            | 0         | 2   | 0     | 0        | 2          |       |
| OC83 (Drivers and mobile plant operators)     | 1         | 0   | 0     | 0        | 1          |       |
| OC96 (Refuse workers and other elementary workers) | 1         | 1   | 1     | 0        | 3          |       |
| NAP†                                         | 35        | 22  | 28    | 36       | 121        |       |
| NRP‡                                         | 1         | 0   | 0     | 1        | 2          |       |
| UNK#                                         | 13        | 11  | 17    | 14       | 55         |       |

†ISCO-08 category not applicable. Mostly students, unemployed, parental leave, etc.
‡ISCO-08 category no response.
§ISCO-08 category unknown. Mostly self-employed, part-time worker, etc., without further specification of the type of work.
managers by Schein (1973). The DI is a questionnaire made up of 92 descriptive terms that were pretested and selected from a set of 131 descriptive terms because ratings on these items differed significantly when they were applied to women versus men (Schein, 1973). Thus, the items used in this study measure attributes stereotypically associated with the male and female gender.

Since the German version of the DI (Schein & Mueller, 1992) was unavailable, we translated the DI from English to German. We used the technique of “back translation.” One co-author translated the DI into German and a student assistant, who was a native English speaker, translated the German version of the DI back into English to check for mismatches. Then we compared their terms with the original expressions. This process resulted in a few minor adjustments. Thus, the 92 items and the instruction were identical in meaning to the English original Descriptive Index.

As in Schein (1973), the ratings on the descriptive items were made on a 5-point scale, ranging from 1 (not characteristic) to 5 (characteristic) with a neutral point of 3 (neither characteristic nor uncharacteristic).

The DI was presented online via Unipark. The materials and data can be downloaded under the link https://osf.io/qz9ur/.

### 2.3 Procedure

In the original study by Schein (1973) the DI was presented in three versions to different participants, one asking for ratings about men, one asking for ratings about women, and one asking for ratings about successful middle managers (Schein, 1973). Since our focus was on stereotypes about scientists in comparison to other stereotypes, we kept the original three conditions used by Schein (1973) but added a fourth condition asking for ratings about successful scientists. Thus, we used a between-subject design with four conditions, differing in the group (i.e., men, women, scientists, or managers) to be rated. Participants received the same instructions as in Schein (1973). We instructed participants to imagine they were about to meet a person for the first time and the only thing known in advance was whether the person was an adult male, female, a successful scientist, or a successful middle-class manager. All four versions had the same descriptive items and instructions, except for the group that participants should rate. We presented the questionnaires in German, which normally uses gendered expressions for managers and scientists when using the typical expression. In this case, one might assume that associations between these groups and men are increased automatically (e.g., Reynolds et al., 2006). To circumvent this effect, we used less common but non-gendered expressions “typische erfolgreiche Person aus dem mittleren Managent” (typical successful person in middle management) for “manager” and “typische erfolgreiche Person aus der Wissenschaft” (typical successful person in science) for “scientist.”

Before participants could take the questionnaire, they were told that participation in this study was completely voluntary and that their data would be saved anonymously on a secure server. They were also informed that there are no health risks to be expected from this study, but that participation may influence their current mood. They were only able to proceed if they fully consented to the conditions of this study (see material at https://osf.io/tqnkh/).

After participants consented by continuing to the next page, they were presented with one of the four instructions corresponding to each of the four conditions. Assignment to the conditions was fully randomized. After the instructions, participants had to rate the descriptive terms from the DI. Descriptive terms were presented with 23 items on each page (four pages total) and a short reminder about the instructions, scale anchors, and the group that was to be rated.

After the DI participants had to provide demographic information about themselves. No additional personal information about the participants was stored.

### 2.4 Power and sensitivity analysis

Following the methods of Schein (1973), we first averaged the data from all participants from each condition for each question and then computed the ICC on the resulting averages. Hence, the degrees of freedom of the test were not determined by the number of participants, but rather by the number of items in the Descriptive Index (see the Results section here and from Schein, 1973, for details why the degrees of freedom only depend on the number of items; see also Duehr & Bono, 2006). Since this number was fixed by the study method, we did not perform a power analysis. Instead, we let the

### Table 2 Demographic distribution of the sample

| Condition | N     | Female | In management | In science | With PhD | Students | Has studied |
|-----------|-------|--------|---------------|------------|----------|----------|-------------|
| Men       | 70    | 47 (67%) | 2 (3%)       | 5 (7%)    | 3 (4%)   | 35 (50%) | 19 (27%)    |
| Women     | 64    | 48 (75%) | 5 (8%)       | 1 (2%)    | 1 (2%)   | 22 (34%) | 29 (45%)    |
| Managers  | 77    | 59 (77%) | 3 (4%)       | 2 (3%)    | 1 (1%)   | 32 (42%) | 30 (39%)    |
| Scientists| 80    | 60 (75%) | 7 (9%)       | 4 (5%)    | 2 (3%)   | 35 (44%) | 33 (41%)    |
| Total     | 291   | 214 (74%) | 17 (6%)     | 12 (4%)   | 7 (2%)   | 124 (43%)| 111 (38%)   |

Note: Percentages in parentheses indicate the proportion of this attribute in each condition.

1All participants who used a term like “Student,” “Studentin,” “Studierender,” etc., as part of the occupation were considered to be a student.

2Participants were considered to have studied if they indicated their highest degree to be at least Bachelor.
survey run for three months and tried to gather as many participants during that time as possible.

While the degrees of freedom of the ICC could not be changed, the accuracy of each average value increases the more participants are collected. Since averages are an unbiased and consistent estimator of the population mean (expected value) if it exists and the variable has finite variance (see Theorem 5.2.6 and Example 10.1.2 by Casella & Berger, 2002), the probability of having a high error decreases the more data is collected; that is for any random variable that has finite variance $\sigma^2$ and for which the expected value $\mu$ exists, the sample mean from $N$ samples is distributed around $\mu$ with variance $\sigma^2/N$ (see Casella & Berger, 2002: Theorem 5.2.6). Unfortunately, to estimate the error for each sample size exactly we would have needed to know the distribution of the values, which was not available. Nevertheless, we tried to estimate the error that is the expected difference to the population mean in a simulation study. For this, we generated 92 random categorical distributions with five categories to represent the unknown distribution on the five possible ratings on each of the items. From these distributions, we drew between $N = 20$ and 100 samples, representing the collected ratings from our participants. The error was estimated by computing the difference between the average and the expected value on each of the 92 items and taking the maximum of all items. This process was repeated 1,000 times for each $N$ to compute the 95% lower interval. The results from this simulation are shown in Figure 1. Our analysis indicates that for the number of 60 participants that were at least included in each condition, no item deviated from the population mean by more than 0.64 with 95% probability. The R-script for this simulation can be found at https://osf.io/mfyp3/.

As part of our exploratory analysis, we also conduct an exploratory factor analysis (EFA). Unfortunately, except for some rules of thumb, no methods to estimate the power or error of an EFA exist (Pearson & Mundform, 2010). The numbers used in these rules of thumb differ widely, ranging from 100 participants to ten times the number of items. 920 participants in our case (Pearson & Mundform, 2010). While our sample size did not meet some of the stricter rules of thumb, we could meet the criteria set for the sample size by at least some of the rules, that is at least 200 participants (fair sample size according to Comrey and Lee, 1992 as cited by Pearson & Mundform, 2010) and three times the number of items (sufficient sample size according to Catell, 1978 as cited by Pearson & Mundform, 2010). Because it cannot be established if the sample size was sufficient based on these rules of thumb alone, we will also present reliability estimates for all factors.

3 | RESULTS

We used R 4.0.0 (R Core Team, 2020) for our analysis with the tidyverse package (Wickham et al., 2019) for data cleaning and restructuring. All raw data and the complete analysis are provided on OSF at https://osf.io/wdjch/.

We will first present the results relating to our hypotheses in the following section. The corresponding analysis employs the methods from Schein (1973) as closely as possible. In the subsequent section, we then provide some more insight into the structure of the detected stereotypes using exploratory factor analysis.

3.1 | Think manager—think male and think scientist—think male effects

We first averaged the ratings for each condition separately for each item, giving us 92 mean ratings for each of the four conditions. Then we computed the intraclass correlations (ICCs) based on these mean ratings. Because Schein (1973) did not indicate which ICC is used, we inferred from the values that the ICC(1,1) must have been used (Shrout & Fleiss, 1979). All tests were performed with $\alpha = 0.05$. A significant ICC indicates that there is some resemblance between the two groups. However, as we expected all groups to show at least some resemblance to each other, we were more interested in the differences between multiple ICCs, that is we did not want to determine if, for example, women show any similarity to managers, but rather if this similarity was stronger or weaker than between other groups. To test if the differences in the similarity between two pairs of groups were significant, we used the methods provided by Duer and Bono (2006; see also Berkery et al., 2013). Thus, the difference between any two ICCs on the DI is significant (at $\alpha = 0.05$) if the correlation differs at least by 0.29.

In Table 3, we present the ICCs for men versus managers, women versus managers, men versus scientists, and women versus scientists. We first analyzed which of the two target groups (scientists and managers) men and women were similar or dissimilar based on the significance of the ICC. Men were significantly similar to both.

![FIGURE 1](image) Expected maximal error on 92 Likert items with 5 points for different numbers of participants.
managers as well as to scientists. Women, however, were only significantly similar to scientists but not to managers (tested at \( \alpha = 0.05 \) without Bonferroni correction). Based on the criterion from Duehr and Bono (2006), men were significantly more similar to both managers and scientists compared to women. These results support Hypothesis 1a and 2a. Because of the low number of male participants (less than 30 male participants in all four conditions), we did not test for any gender differences. Appendix A shows that a similar pattern was found if only female participants were included in the analysis, but the strength of stereotypes was somewhat reduced. That is if men were removed from the analysis all four groups became more similar to each other.

To understand which of the items caused these differences, we performed a Tukey HSD test on each of the 92 items separately. The Tukey HSD is similar to the Duncan multiple range test (DMRT) for unequal \( n \), which was used by Schein (1973). Jaccard et al. (1984) recommended the Tukey HSD over the DMRT due to its more desirable statistical properties. Because of the 92 simultaneous tests, we used a Bonferroni correction to assess which differences were significant. We found that women differed significantly from managers as well as from scientists by 54 of the 92 items (59%). Men, however, only differed from managers on 17 items (18%) and 30 items from scientists (33%). To test if these differences were significant, we used a McNemar test with continuity correction on the number of items that differed significantly for men versus women from scientists and managers. This test indicated that women differed both from managers on more items (\( \chi^2 (1) = 21.25 \), two-tailed \( p < .01 \)) as well as from scientists on more items (\( \chi^2 (1) = 11.02 \), two-tailed \( p < .01 \)) than men. This supports Hypotheses 1b and 2b.

### 3.2 Exploratory analysis of the stereotype structure

We visualized the similarities between the four groups, by translating the correlations into distances using the formula \( d = \sqrt{2(1-r)} \) and then computing a multidimensional scaling with two dimensions on the resulting distance matrix. The resulting solution is presented in Figure 2. The multidimensional scaling revealed that participants rated both managers as well as scientists as similar professions, supporting our notion that both were regarded as leadership positions. Men were closer to both scientists as well as to managers than women. Women seemed to be isolated from all other groups, with the largest distance to all of them, showing that stereotypes about women differ from stereotypes about men, managers, as well as scientists.

To investigate on which attributes the four groups differed most, we performed an exploratory factor analysis (EFA) on all responses. To this end, we first used a PCA with a parallel analysis (PA, Horn, 1965; see also Hayton et al., 2004) to estimate the number of factors. A PA compares the variance explained of the PCA components on the collected data with those on random data to test which of the components provide a variance explanation above chance level. The PA method outperforms many other methods that can be used to detect the number of factors in a data set (Hayton et al., 2004) and is considered the gold standard to select the number of factors in an EFA by some (Glorfeld, 1995). To generate random data sets for the parallel analysis, we bootstrapped new
random answers to each item independently for each item. We ran 1,000 repetitions of the parallel analysis (generation of random data set and PCA), to ensure that the estimated variances were representative. Figure 3 shows the scree plot from the PCA with the results from the PA for reference.

Because the DI is based on items that differed most between men and women (two theoretical factors; Schein, 1973) and previous studies only considered a small number of factors in their analysis of the DI (Duehr and Bono (2006): three factors; Carli et al. (2016): four factors), we chose to be conservative and selected as few factors as possible. Therefore, we based the number of factors on the upper limit of the 95% CI of the PA and retained factors for which the variance explained was higher than this number (Hayton et al., 2004). A six-factor solution provided the highest variance explanation (cumulative variance: 51.5%, proportion of variance from the sixth PCA component: 2.7% versus 2.2% [2.1%; 2.3%] for random data) without overfitting (proportion of variance from the seventh PCA component: 2.1% versus 2.1% [2.0%; 2.2%] for random data). For comparison, Bartlett’s chi-square test (Bartlett, 1950) indicated that seven factors should be extracted ($\chi^2 = 3496.27, p = .78$ for seven factors versus $\chi^2 = 3996.73, p < .01$ for six factors). Because Bartlett’s chi-square test has been shown to over-extract sometimes (Hayton et al., 2004), we decided to retain only six factors as indicated by the PA. We then used a factor analysis with a Promax rotation (R package psych, Revelle, 2019) to extract six factors. These six factors provided an off-diagonal fit of 0.97 (for comparison: a two-factor solution had a fit of 0.86, a three-factor solution like the one from Duehr & Bono, 2006 had a fit of 0.93, and a four-factor solution like the one from Carli et al., 2016 had a fit of 0.95). We used the resulting loading matrix to generate scales by assigning each item to the scale on which it had a loading higher than 0.39. If an item loaded highly on multiple factors or did not load sufficiently on any of the six factors, the item was dropped from the scales (14 items were dropped). We chose the cutoff value of 0.39 because this value allowed us to retain the maximum number of items. Based on the remaining items, values for the scales were calculated by z-transforming the answers and then computing the mean for all factors.

z-scores from the same scale, with reversed z-scores for items that had negative loadings on the corresponding factors.

By inspecting the items contained in each scale, we assigned names to the scales that summarized the items in each scale. We also compared these scales to the leadership styles from Duehr and Bono (2006; Agentic, Communal, Task-Oriented, Relationship-oriented) and the four scales of positive and negative agentic and communal from Carli et al. (2016; Agentic; Communal; Dissociation, i.e., negative communal; Passive, i.e., negative agentic). The six resulting scales were: leading, which mostly corresponded to the agentic (4 items) and task-oriented (4 items) scales used by Duehr and Bono (2006) as well as to the agentic scale (15 items) used by Carli et al. (2016); social, which mostly corresponded to the communal (4 items) and relationship-oriented (3 items) scale used by Duehr and Bono (2006) as well as to the communal scale (12 items) used by Carli et al. (2016); academic, which mostly corresponded to the task-oriented (4 items) and agentic (2 items) scales used by Duehr and Bono (2006) as well as to the agentic scale (9 items) used by Carli et al. (2016); anti-social, which did not correspond to any of the scales used by Duehr and Bono (2006; 1 item from the agentic scale and 1 item from the communal scale) but corresponded to the passive (10 items) and dissociative scale (3 items) used by Carli et al. (2016); compliant, which did not contain any of the items categorized by Duehr and Bono (2006) but corresponded to the communal scale (2 items) used by Carli et al. (2016); and rebellious, which did not correspond to any of the scales used by Duehr and Bono (2006; 1 item from the task-oriented scale) but corresponded to the agentic scale (4 items) used by Carli et al. (2016) with one additional reversed item from the communal scale. Another comparison may be made with the ABC of stereotypes (Koch et al., 2016), which indicate that stereotyped groups are not primarily judged based on their communal attributes, but rather by their agency/success, and their beliefs (conservative versus progressive). While the ABC of stereotypes also includes communion as a stereotypical attribute, high or low communion is only attributed to a group if the agency is high enough to fulfill communal or uncommunal role. In this comparison, one might interpret the anti-social scale as a scale that purely indicates uncommunal tendencies, whereas the rebellious scale corresponds more directly to progressive beliefs from the ABC. Similarly, the compliant scale would indicate conservative beliefs in combination with a communal attitude. The full set of items used in each scale is given in Table 4. We also estimated the reliability of the resulting factors by estimating Cronbach’s $\alpha$. For four of the six resulting factors, the reliability was sufficient ($\alpha > 0.70$; Nunally & Bernstein, 1994), whereas the last two factors only had low reliability. The most likely reason for the low reliability of these factors is the low number of items contributing to those factors, showing that the DI is biased toward the other four factors and disregards differences among the last two. Due to the exploratory nature of this analysis, we nevertheless decided to keep these two factors despite their low reliability. The resulting factors did not depend significantly on the age of the participants ($p > .12; |r| < 0.07$ for all factors based on a Kendall rank correlation with Bonferroni correction for six simultaneous tests). There was a
| Factor Leading | Social | Academic | Anti-Social | Compliant | Rebellious |
|---------------|--------|----------|-------------|-----------|------------|
| High need for power | Sympathetic (C,c) | Consistent (a) | Uncertain (p) | interested in own appearance | Adventurous (a) |
| Leadership ability (a) | values pleasant surroundings | Neat | Devious (d) | Talkative (c) | Frank (a) |
| Emotionally stable | Courteous (c) | Intelligent (T,a) | Frivolous (p) | Not conceited about appearance* | Independent (T,a) |
| Forceful (A,a)¹ | Desire for friendship (c) | Persistent (a) | Timid (p) | Strong need for social acceptance (c) | Strong need for security* (c) |
| Competitive (T,a) | Cheerful (c) | Sophisticated | waver in decision (p) | High need for autonomy (a) |
| Able to separate feelings from ideas (a) | Understanding (R) | Analytical ability (A,a) | Vulgar | Obedient* |
| High self-regard (a) | Grateful (c) | Competent (T) | Aggressive (A) | |
| Speedy recovery from emotional disturbance (T) | Aware of feelings of others (C,c) | Industrious (T,a) | Easily influenced (p) | |
| Firm (a) | Humanitarian values (c) | well informed (a) | Exhibitionist | |
| Prompt (a) | Modest | Ambitious (A,a) | Dawdler and procrastinator (p) | |
| Desires responsibility (a) | Sentimental (c) | Logical (T,a) | Quarrelsome | |
| Self-controlled | Tactful (R,c) | Knows the way of the world (a) | Reserved (p) | |
| Decisive (T,a) | Helpful (C,c) | Hasty (p) | |
| Direct (a) | Generous (c) | Nervous (p) | |
| Hides emotion | Kind (C,c) | Deceitful (d) | |
| Authoritative (a) | Intuitive (R) | Bitter (d) | |
| Self-confident (A,a) | | Submissive (C,p) | |
| Assertive (A,a) | | | |
| Feelings not easily hurt (a) | | | |
| Dominant (A,a) | | | |
| Skilled in business matters (T) | | | |
| Strong need for monetary rewards | | | |
| Self-reliant (a) | (Steady) | (Curious) | (Fearful) | |
| | (Desire to avoid controversy) | | | |
| | (Strong need for achievement) | (Vigorous) | (creative) | (Passive) |
| | (Demure) | (sociable) | (Objective) | (Shy) |
| No Items | 23 | 16 | 12 | 17 | 4 | 6 |
| %Var | 13.9 | 10.4 | 9.7 | 9.7 | 4.6 | 4.7 |

Note: Letters in parentheses indicate the scales used by Duehr and Bono (2006; A: Agentic, C: Communal, T: Task-Oriented, R: Relationship-oriented) and Carli et al. (2016, a: Agentic, c: Communal, d: Dissociation, p: Passive).

Notes Items in parentheses indicate items that were dropped, because they did not load sufficiently on any of the items or because they loaded on multiple items. Variance explained is calculated on the complete loading matrix without dropped items.

¹Changed to dominant by Carli et al. (2016).
significant gender difference in the ratings of the factor academic showing that male participants generally gave lower scores on this factor (β = −0.33; SE = 0.09; t(289) = −3.72; p < .01) and for the factor anti-social showing that male participants generally gave higher scores on this factor (β = 0.28; SE = 0.07; t(289) = 3.76; p < .01). All other factors were not influenced significantly after Bonferroni correction (all |β| < 0.24; |t(289)| < 2.72; p > .007).

Based on these scales, we used a Tukey HSD test with Bonferroni corrections for six simultaneous tests to estimate which of the groups differed significantly. Also, we used a two-sided one-sample t test with Bonferroni corrections for 24 simultaneous tests (6 scales x 4 groups) to test which group means differed significantly from zero (the group-mean resulting from the z-transformation). To analyze if there are any age differences in the stereotypes on specific groups, we calculated Kendall rank correlations of the factor scores and the age of the participants separately for each group. None of the correlations showed a significant influence of age on the ratings after Bonferroni correction for six simultaneous tests (|r| < .19; p > .03). Thus, there was no age difference in the structure of the stereotypes. All results are summarized in Table 5 and visualized in Figure 4.

### 3.2.1 | Leadership scale

These tests showed that both men and managers were seen as above average in leading, whereas women were seen as below average. Scientists were seen as average in terms of leadership. Also, women differed significantly from both men and scientists in their associated leadership attributes. This indicated that the female gender was associated with a lack of leadership attributes required both for management positions as well as for positions as scientists.

### 3.2.2 | Social scale

For social attributes, women differed significantly from all other three groups. Women had significantly higher values compared both to the overall mean as well to all other three groups. Men, managers, and scientists did not differ significantly from each other. However, both managers, as well as men, were seen as significantly below average, whereas scientists were seen as average.

### 3.2.3 | Academic scale

Academic attributes were similarly low for both genders, but higher for both scientists and managers. Unsurprisingly, academic attributes were more strongly associated with scientists. Managers were seen as average in terms of academic attributes.

### 3.2.4 | Anti-social scale

Similarly, for anti-social attributes, there was no significant gender difference, with both genders having descriptively higher scores on the anti-social scale than either professional group. However, only men differed significantly from the overall sample mean. Both scientists and managers were associated significantly less with anti-social attributes than men or women. However, only for scientists, the scores on the anti-social scale were significantly below the overall mean.

### 3.2.5 | Compliance scale

Compliance was significantly lower than average for both men and scientists with no significant difference between these two groups, but significantly higher for managers and women. However, because women were much more strongly associated with compliant attributes than managers, the difference between managers and women was also significant.

### 3.2.6 | Rebellious scale

Finally, rebellious attributes were descriptively associated more strongly with men and scientists, with only men differing significantly

---

**TABLE 5** Mean ratings of the four groups on the six extracted factors

| Factor        | Men         | Women       | Managers    | Scientists   | % Var | α               |
|---------------|-------------|-------------|-------------|--------------|-------|-----------------|
| Leading       | 0.24 ± 0.08 | −0.64 ± 0.47| 0.29 ± 0.51 | 0.01 ± 0.22  | 9.5   | 0.93 ± 0.90 ± 0.95 |
| Social        | −0.20 ± 0.36| 0.75 ± 0.94 | −0.27 ± 0.03| −0.16 ± 0.05 | 7.7   | 0.92 ± 0.89 ± 0.94 |
| Academic      | −0.34 ± 0.51| −0.41 ± 0.36| 0.11 ± 0.36 | 0.52 ± 0.77  | 6.1   | 0.90 ± 0.87 ± 0.92 |
| Anti-Social   | 0.29 ± 0.46 | 0.19 ± 0.37 | −0.17 ± 0.08| −0.25 ± 0.05 | 6.0   | 0.87 ± 0.84 ± 0.90 |
| Compliant     | −0.30 ± 0.50| 0.59 ± 0.80 | 0.22 ± 0.42 | −0.42 ± 0.17 | 2.1   | 0.12 ± 0.09 ± 0.30 |
| Rebellious    | 0.40 ± 0.54 | −0.42 ± 0.21| −0.14 ± 0.12| 0.12 ± 0.36  | 2.7   | 0.25 ± 0.07 ± 0.41 |

Note: All confidence intervals are Bonferroni corrected to achieve 95% simultaneous confidence. Groups in each row that do not share a subscript differed significantly based on a Tukey HSD test with α = 0.008 (equivalent to α = 0.05 after Bonferroni correction for six simultaneous tests). % Var indicates the variance explanation of the scales after ambivalent items were dropped and all items were included with equal weight in the scale. *p < .002 (equivalent to p < .05 after Bonferroni correction for 24 simultaneous tests) that the value is not equal to zero (sample mean) based on a two-sided t test.
from the sample mean and no significant difference between men and scientists. Women were rated as least rebellious, with no significant differences between women and managers, but only women having rebellion scores significantly below the overall mean.

4 | DISCUSSION

The current study tested whether the think manager—think-male effect and the think scientist—think-male effects are driven by the same attributes being associated with leadership. We investigated if the think manager—think male effect is still present after being observed over 25 years ago in Germany. Furthermore, we examined the attributes associated with scientists and their relation to attributes associated with women and men. Women were not only rated as less similar to managers compared to men but also scientists. Thus, in addition to a think manager—think-male effect, we also found a think scientist—think male effect, successfully replicating Carli et al. (2016) in a German sample. However, the think scientist—think male effect was smaller than the think manager—think male effect. Most importantly, when we investigated the structure of these two effects, we found that these effects were driven by different attributes. We found that men were perceived as more similar to managers than to women because both men and managers were seen as strong in leadership or agentic attributes and were less associated with social attributes. Women were only seen as similar to managers in that both groups were seen as less rebellious than average. These associated attributes contrast with other findings, where researchers found that leadership roles are changing toward a more androgynous role (Koenig et al., 2011; Lueptow et al., 2001; but see also Berkery et al. 2013 for some findings contradicting this trend).

4.1 | Comparison of the stereotype profiles

Unlike in previous studies (e.g., Carli et al., 2016), we did not group the items of the Schein DI into agentic and communal traits, but created a more detailed factor structure, resulting in the six factors leading, social, academic, anti-social, compliant, and rebellious. While these scales are currently only exploratory and should be validated further, a more detailed view of the stereotype structure can provide more insight on why different groups are perceived differently here and in future research. Based on these scales we can postulate some stereotype profiles for each of the four groups.

At first glance, this could indicate a common underlying component of the think manager—think male and the think scientist—think male effect. However, scientists were only associated with average leadership attributes and significantly less with leadership attributes than managers, indicating a different perception of managers and scientists in terms of their leadership attributes. Unlike with women and managers, the mismatch between women and scientists did not arise from scientists perceived as strong leaders. Quite the opposite, typical scientists were associated less with leadership attributes than typical men, although this tendency was not significant. However, women were seen as far below average and this caused a significant difference between women and scientists.

Men are seen as leaders, but they aggressively (anti-social) work against existing norms, both by questioning these norms (rebellious) as well as by ignoring them (lack of compliant attributes). Women are mostly defined by their communal attributes and as followers rather than leaders. They are seen as following social norms, instead of questioning or working against them. Managers are seen as leaders who follow social norms (compliant). Scientists are mostly defined by their academic status. They disregard social norms (lack of compliant attributes), but they lack the aggressive tendencies against social norms typical of men (lack of anti-social attributes). In terms of beliefs, scientists are seen as rather progressive. This might be due to the role of scientists, whose job it is to bring forth innovative ideas not covered by existing social norms or knowledge.

Since discrimination works both through descriptive as well as prescriptive stereotypes (Eagly & Karau, 2002), it is important to see which of these assigned scales relates to either category. Since social stereotypes about women are mostly prescriptive (Eagly & Karau, 2002), a violation of these prescribed norms can arise if women take positions as managers or scientists, both of which are associated less with social attributes. Most importantly, the finding that scientists are not seen as particular social was in line with previous research on science as lacking prospects for the attainment of communal goals (Diekman et al., 2010). This provides a strong hint that discrimination for women in academic and management positions works through penalties for perceived norm violations as well as self-selection since potential communal goals are believed to be
unattainable through a career in science. This indicates an influence of the role stereotype of scientists, rather than the stereotype about women.

Similarly, the association of women with compliant attributes indicates that norm violations might occur, if they attempt to choose a career as a scientist. At the same time, no such norm violations occur in terms of compliance, if women take the career path of a manager, as women and managers did not differ as much in terms of these attributes. The same is true for rebellious attributes, which were more associated with men and scientists than with women. Moreover, we found that men were associated significantly less with compliant and significantly more with rebellious attributes. It is not clear if this is a descriptive or prescriptive stereotype about men. One could assume that this part of a tendency to see men as freedom seekers, wild, etc. As men and scientists are comparable in their perceived lack of compliance and their rebellious attributes, this could create a situation like the sticky-floor effect in which men are driven toward the role of scientists. Thus, such a prescriptive stereotype could also explain the lack of women in scientific positions, as they do not share this external drive toward these types of occupations. In this case, there would be two forces creating discrimination when women attempt a career as scientists. Discrimination due to norm violations when women try to fulfill the less compliant role of scientists, as well as the pull for men to take such positions that allow them to fulfill the roles assigned by society.

The lack of difference in academic attributes between men and women suggests that discrimination is not so much driven by the belief that women are unable to fulfill the role of a successful scientist, but rather the role of a scientist is believed not to be fitting to women.

While the ICC between men and scientists was smaller than between men and managers, men overlapped in four of the identified factors with scientists compared to only two factors with managers. This leads to somewhat ambiguous results because it is not clear if the stereotype of men coincides more with managers or more with scientists. However, this implies a clear difference in the type of discrimination that women might face when approaching a career as a manager and when approaching a career as a scientist. Whereas a large part of the discrimination against women approaching a career as a manager comes from the fact that women are seen as unable or unwilling to lead, for scientists, the mismatch arises due to the strong tendency of stereotypical women to follow norms instead of working against them, whereas scientists are seen as frequently violating these norms.

4.2 | Practical implications

Just like the think manager—think-male effect, the think scientist—think-male effect arose partially from scientists/managers and men being seen as more similar concerning leadership and social attributes. Both scientists, as well as managers, are, therefore, seen as positions requiring some degree of leadership, and the corresponding leadership attributes were seen as more present in men. However, as Duehr and Bono (2006) found, this is specific to the task-oriented leadership style that is seen to be less fitting for women, at least in their student sample. Relationship and transformational leadership styles are instead seen as more fitting for women than for men. Because of the survey we used, we cannot assess how strongly transformational leadership styles are associated with scientists. However, since a large part of the work of scientists is in educating junior researchers, a transformational leadership style is beneficial in this position (Balwant, 2016). Increasing the perception of scientists as teachers of junior scientists and students in addition to their other roles can, therefore, help reduce the stereotypes that hinder women from being seen as successful scientists. For example, while the works of famous scientists are often known, it is much less clear which successful scientists taught other scientists (see also Clark et al., 2016; Ramsey, 2020).

While the similarities in leadership attributes between men and scientists caused some overlap of the stereotypes, scientists were nevertheless perceived as having fewer leadership attributes compared to typical men, although this difference was not significant. Moreover, scientists were significantly less associated with leadership attributes compared to managers. Thus, differences in leadership attributes can only explain a small portion of the think scientist—think male effect. A stronger difference between women and scientists could be found among the communal attributes. Scientists differed from women on the social, compliant, and rebellious factor, whereas men were seen as similar to scientists on all three of these factors. The difference was most striking for the social factors, with men, managers, and scientists all being seen as similar, and women being seen as having significantly more social attributes. Many of the items in the social scale are related to helping or supporting others and this tendency is seen more strongly in women than in any of the other groups. As previous studies have indicated, these stereotypes do not only relate to how women are seen, but also to the norms applied to women and the goals they want to achieve (Eagly & Karau, 2002). As Clark et al. (2016) have shown, perception of science as not serving communal goals can hinder women from entering science careers. We confirm that scientists are not seen as particularly communal and also may be seen as disregarding social norms. Thus, women pursuing communal goals may be deterred from following scientific careers due to a goal conflict. Highlighting the communal achievements of scientists and therefore changing the role stereotype may then both help women with those types of goals to consider a scientific career (Clark et al., 2016), as well as reduce the perceived lack of fit between women and scientists. Similarly, topics from natural sciences are seen as more scientific than topics from the behavioral and social sciences (Krull & Silvera, 2013). This may also contribute to science being seen as not particularly social. As Krull and Silvera (2013) point out, however, science is defined by the methods and not by the topics. Thus, changing the public perceptions of science by providing more examples of social topics
researched scientifically may also change the image of scientists and increase the perception of scientists as communal.

Likewise, scientists were seen as norm-breakers with more progressive beliefs than the other groups (i.e., lowest scores on the compliant scale). This role was not seen as fitting to women, who were seen as the most compliant. However, it is noteworthy that most items on the compliance scale only corresponded to issues of appearance, and did not indicate any actual uncommunal tendencies. Therefore, it might be helpful to highlight situations where scientists have benefitted humanity by ignoring or questioning existing conservative beliefs to show that these progressive beliefs are compatible with an overall communal tendency in science.

Also, our results suggest that interventions highlighting similarities in academic abilities of men and women will only have little influence on discrimination. Unlike other studies (e.g., Nosek et al. 2002; Nett et al., 2020) we found no difference between men and women in terms of the associated academic attributes. Thus, our findings are in line with studies that show women to be perceived at least as competent as men (Eagly et al., 2020). This indicates that discrimination in science results from other factors than perceived differences of competence. Most importantly, our research shows that while the think manager—think male effect is mostly driven by stereotypes about women, the think scientist—think male effect is much more driven by stereotypes about scientists (see Appendix B). This also explains why increased exposure was already able to reduce this effect (e.g., Miller et al. 2015, 2018). In combination with the proposed self-selection mechanisms, this indicates that further changes in the stereotypes about scientists may increase gender diversity in science.

4.3 | Limitations

Although the Schein DI was created based on stereotypes about men and women (Schein, 1973), it is heavily biased toward measuring leadership attributes (23 out of the 92 items), and mostly focuses on agentic and task-oriented leadership styles (Duehr & Bono, 2006). It is more sensitive to overlap in leadership attributes than to an overlap in the other factors. Additionally, the Schein DI is lacking some attributes that specifically correspond to stereotypes about scientists (Carli et al., 2016). Nevertheless, to make our results comparable to the original results provided by Schein (1975, 1975) we chose to use the original version of the Schein DI instead of one of the updated versions. However, this may have biased the results and prevented us to get a complete picture of the stereotypes about the groups used (cf. Koch et al., 2016). Furthermore, we currently cannot distinguish if the overlap between men and scientists is weaker (lower ICC) or stronger (more matching dimensions) than between men and managers. Also, the Schein DI was composed of items that were elicited almost 50 years ago (Schein, 1975). Since stereotypes can change, this may have led to some items not being representative of stereotypes of men and women (cf. Devine & Elliot, 1995). Moreover, the Schein DI was compiled from items that differed most between men and women in a pretest (Schein, 1973). Thus, the Schein DI is much more sensitive to differences between the genders than to similarities. Because of this, differences measured through the Schein DI may not reflect the actual perception of men, women, managers, or scientists. Most importantly, it may not be able to fully capture the stereotypes about scientists or managers. Thus, if these stereotypes also play a role in addition to stereotypes about men and women, then the DI may not be able to provide a complete picture of the reasons for the discrimination. Thus, future research should try to extend the DI to provide also a detailed picture about stereotypes for other groups that may play a role, instead of just relying on stereotypes about men or women. Such an approach has already been undertaken by Duehr and Bono (2006) and Carli et al. (2016) who extended the DI with attributes corresponding to different types of managers or scientists, respectively. Instead of focusing on attributes that differ most between men and women, a true measure of stereotypes should focus on attributes that come to mind spontaneously when talking about a group and therefore represent effortlessly and habitually activated assumptions about members of that group (cf. Koch et al. 2016). Another limitation arises from the fact that we collected a convenience rather than a representative sample of the population. We explicitly decided to not use a representative sample to be able to collect as many participants as possible. However, this makes it less clear if the results we found can be generalized to the whole population, or if these findings only apply to a smaller group within this population.

5 | CONCLUSION

In conclusion, we found evidence that stereotypes may not only hinder the careers of women if they approach management positions but also in academia. While the think manager—think male effect has been found to become reduced over time, this may partially be attributed to different stereotypes about agentic traits of managers (Duehr & Bono, 2006; but see also Haines et al., 2016 for alternative explanations). However, for the think scientist—think male effect, stereotypes about scientists did not include leadership attributes, but we still found a mismatch between scientists and women. Therefore, the same mechanism that leads to a reduction of the think manager—think male effect cannot account for this. Instead, the think scientist—think male effect was mainly driven by the ratings of women as more social compared to the other groups. Thus, while women in science can partially gain from a change in the leadership roles in science, a stronger effect can be expected if communal goals of science are more strongly highlighted.

PARTICIPANT CONSENT

Consent for taking part in the study was given by all involved subjects as part of the survey.
CONFLICT OF INTEREST
None of the authors have any conflict of interest to declare.

DATA AVAILABILITY STATEMENT
Data corresponding to this study was uploaded and registered on OSF at https://osf.io/csn4w DOI: 10.17605/OSF.IO/CSN4W.

ORCID
Nadine Nett https://orcid.org/0000-0002-0856-7146
Julia Englert https://orcid.org/0000-0002-5974-9265

REFERENCES
Adler, N. J. (2002). Global managers: No longer men alone. International Journal of Human Resource Management, 13(5), 743–760. https://doi.org/10.1080/095851902125895
Bakan, D. (1966). The duality of human existence: An essay on psychology and religion. Rand McNally.
Balwant, P. T. (2016). Transformational leadership in higher education teaching: A meta-analytic review and research agenda. Journal of Leadership Studies, 9(4), 20–42. https://doi.org/10.1002/jls.21423
Bartlett, M. S. (1950). Tests of significance in factor analysis. British Journal of Statistical Psychology, 3(2), 77–85. http://dx.doi.org/10.1111/j.2044-8317.1950.tb00285.x
Bench, S. W., Lench, H. C., Liew, J., Miner, K., & Flores, S. A. (2015). Gender gaps in overestimation of math performance. Sex Roles, 72(11), 536–546. https://doi.org/10.1007/s11199-015-0486-9
Berkery, E., Morley, M., & Tiernan, S. (2013). Beyond gender role stereotyping: A cross cultural analysis. Journal of Applied Social Psychology, 43(8), 711–721. https://doi.org/10.1177/0021902913500494
Booysen, L. A., & Nkomo, S. M. (2010). Gender role stereotypes and requisite management characteristics: The case of South Africa. Gender in Management: An International Journal, 25(4), 285–300. https://doi.org/10.1108/17542411011048164
Braun, S., Stegmann, S., Bark, A. S. H., Junker, N. M., & van Dick, R. (2017). Think manager—think male, think follower—think female: Gender bias in implicit followership theories. Journal of Applied Social Psychology, 47(7), 377–388. https://doi.org/10.1111/jasp.12445
Carli, L. L., Alawa, L., Lee, Y., Zhao, B., & Kim, E. (2016). Stereotypes about gender and science: Women ≠ scientists. Psychology of Women Quarterly, 40(2), 244–260. https://doi.org/10.1177/0361684315622645.
Casella, G., & Berger, R. L. (2002). Statistical inference (Vol. 2). Pacific Grove, CA: Duxbury.
Ceci, S. J., & Williams, W. M. (2011). Understanding current causes of women’s underrepresentation in science. Proceedings of the National Academy of Sciences, 108(8), 3157–3162. https://doi.org/10.1073/pnas.1014871108
Charlesworth, T. E., & Banaji, M. R. (2019). Gender in science, technology, engineering, and mathematics: Issues, causes, solutions. Journal of Neuroscience, 39(37), 7228–7243. https://doi.org/10.1523/JNEUROSCI.0475-18.2019
Clark, E. K., Fuesting, M. A., & Diekman, A. B. (2016). Enhancing interest in science: Exemplars as cues to communal affordances of science. Journal of Applied Social Psychology, 46(11), 641–654. https://doi.org/10.1111/jasp.12392
de Pillis, E., Kernochan, R., Melich, O., Prosser, E., & Whiting, V. (2008). Are managerial gender stereotypes universal? The case of Hawai’i. Cross Cultural Management: An International Journal, 15(1), 94–102. https://doi.org/10.1108/13527600810848854
Devine, P. G., & Elliot, A. J. (1995). Are racial stereotypes really fading? The Princeton trilogy revisited. Personality and Social Psychology Bulletin, 21, 1139–1150. https://doi.org/10.1177/0146167295211002
Diekman, A. B., Brown, E. R., Johnston, A. M., & Clark, E. K. (2010). Seeking congruity between goals and roles: A new look at why women opt out of science, technology, engineering, and mathematics careers. Psychological Science, 21(8), 1051–1057. https://doi.org/10.1177/0956797610377342
Diekman, A. B., & Eagly, A. H. (2000). Stereotypes as dynamic constructs: Women and men of the past, present, and future. Personality and Social Psychology Bulletin, 26(10), 1171–1188. https://doi.org/10.1177/0146167200262001
Duehr, E. E., & Bono, J. E. (2006). Men, women, and managers: Are stereotypes finally changing? Personnel Psychology, 59(4), 815–846. https://doi.org/10.1111/j.1744-6570.2006.00055.x
Dunn, A. M., Heggestad, E. D., Shanock, L. R., & Theilgaard, N. (2018). Intra-individual response variability as an indicator of insufficient effort responding: Comparison to other indicators and relationships with individual differences. Journal of Business and Psychology, 33(1), 105–121. https://doi.org/10.1007/s10869-016-9479-0
Eagly, A. H. (2013). Sex differences in social behavior: A social-role interpretation. Psychology Press.
Eagly, A. H., & Johannesen-Schmidt, M. C. (2001). The leadership styles of men and women. Journal of Social Issues, 57(4), 781–797. https://doi.org/10.1111/0022-4537.00241
Eagly, A. H., & Karau, S. J. (2002). Role congruity theory of prejudice toward female leaders. Psychological Review, 109(3), 573–598. https://doi.org/10.1033/0033-295X.109.3.573
Eagly, A. H., Nater, C., Miller, D. I., Kaufmann, M., & Szczesny, S. (2020). Gender stereotypes have changed: A cross-temporal meta-analysis of U.S. public opinion polls from 1946 to 2018. American Psychologist, 75(3), 301–315. https://doi.org/10.1037/amp0000494
European Commission. (2015). She Figures 2015. https://ec.europa.eu/research/swafs/pdf/pub_gender_equality/she_figures_2015-final.pdf
European Commission. (2017). 2017 Report on equality between women and men in the EU. https://eaeu.eea.europa.eu/sites/eeas/files/2017_reportequality_women_men_in_the_eu_en.pdf
Eurostat. (2018). Gender statistics. https://ec.europa.eu/eurostat/estat-explained/index.php/Gender_statistics
Finkelstein, M. J., Conley, V. M., & Schuster, J. H. (2016). Taking the measure of faculty diversity. TIAA Institute.
Fiske, S. T., Cuddy, A. J. C., Glick, P., & Xu, J. (2002). A model of (often mixed) stereotype content: Competence and warmth respectively follow from perceived status and competition. Journal of Personality and Social Psychology, 82(6), 878–902. https://doi.org/10.1037/0022-3514.82.6.878
Fullagar, C. J., Sumer, H. C., Sverke, M., & Slick, R. (2003). Managerial gender stereotypes finally changing? Personnel Psychology, 56(4), 285–300. https://doi.org/10.1177/0956797602105878
Glöckner, A., & Betsch, T. (2011). The empirical content of theories in judgment and decision making: Shortcomings and remedies. Judgment and Decision Making, 6(8), 711–721. https://doi.org/10.1177/205525461142031104
Grönlund, S., & Liljedahl, U. (2011). An improvement on Horn’s parallel analysis methodology for selecting the correct number of factors to retain. Educational and Psychological Measurement, 71(3), 377–393. https://doi.org/10.1177/0013164910386244
Griffiths, O., Roberts, L., & Price, J. (2019). Desirable leadership attributes are preferentially associated with women: A quantitative study of gender and leadership roles in the Australian workforce. Australian Journal of Management, 44(1), 32–49. https://doi.org/10.1077/0312896218781933
Gruber, J., Mendle, J., Lindquist, K. A., Schnader, T., Clark, L. A., Bliss-Moreau, E., Akinola, M., Atlas, L., Barch, D. M., Barrett, L. F., Borelli,
APPENDIX A

Analysis of female participants only
When only female participants were included, men were significantly similar to both managers ($r = 0.71 [0.59, 0.79]$; $F(92,91) = 5.78 [3.82, 8.73]$; $p < .01$) as well as to scientists ($r = 0.58 [0.42, 0.70]$; $F(92,91) = 3.73 [2.47, 5.63]$; $p < .01$) as in the original analysis. Unlike in the original analysis, however, women, were also significantly similar to managers ($r = 0.19 [-0.02, 0.38]$; $F(92,91) = 1.46 [0.97, 2.20]$; $p = .04$) as well as to scientists ($r = 0.28 [0.08, 0.46]$; $F(91,91) = 1.77 [1.17, 2.68]$; $p < .01$). Comparing the ICCs we found that like in the original analysis men were significantly more similar to both managers and scientists than women. The Tuckey-HSD test for female participants showed that women differed from managers on 50 of the 92 items (54 items in the original analysis) and scientists on 46 items (54 items in the original analysis). Men differed from managers on 13 of the 92 items (17 in the original analysis) and scientists on 20 of the 92 items (30 in the original analysis). A McNemar test with continuity correction confirmed that men and women differed significantly in the number of items that were similar to managers ($X^2 (1) = 24.45; p < .01$) and in the number of items that were similar to scientists ($X^2 (1) = 13.59; p < .01$).

APPENDIX B

Test for Group versus Role stereotypes
In our exploratory analysis we found that the think-scientist–think male effect may be driven by the stereotype about scientists as well as the stereotype about women, whereas the think-manager–think male effect is mainly driven by the stereotype about women. Here we propose an analysis method to distinguish between these two types of effects. Since correlations can be interpreted as distances, it is possible to compute the three pairwise distances between men, women, and the third group. These three distances uniquely define a triangle with the respective side lengths. If the third group is located along the imaginary men-women line, this triangle will have a flat angle at this point. However, if this point is located at a different position, then not just an interpolation from typical male/female features, then the angle will be more acute, and this group will have a distinct stereotype that is neither typically male nor female (see Figure 2 for comparison). Thus, the angle in this imaginary triangle can be taken as a measurement, angles close to 180° indicate that the effect is driven by the male-female stereotypes, whereas lower angles indicate that the effect is driven by stereotypes of the third group. Based on the three lengths, the angle may be computed using the cosine theorem $\alpha = \cos^{-1} \left( \frac{a^2 + b^2 - c^2}{2bc} \right)$ where $a$ is the distance between men and women. We used this method on our data with bootstrapping (10,000 bootstraps) and found that the angle for managers was descriptively higher ($\alpha = 93.9^\circ [89.2^\circ; 98.4^\circ]$) than the angle for scientists ($\alpha = 88.0^\circ [84.2^\circ; 91.7^\circ]$), indicating that the think-scientists–think male effect is driven by stereotypes about scientists to a higher degree than the think-manager, think-male effect. We propose that future research on this topic should not only report the ICCs but also this angle so that changes in the stereotypes can be better understood.

How to cite this article: Nett N, Nett T, Englert J, Gaschler R. Think scientists—Think male: Science and leadership are still more strongly associated with men than with women in Germany. J Appl Soc Psychol. 2021;00:1–17. https://doi.org/10.1111/jasp.12761