Social consensus influences ethnic diversity preferences

Fredrik Jansson and Moa Bursell

ABSTRACT
There is widespread segregation between workplaces along ethnic lines. We expand upon previous research on segregation and social influence by testing the effect of the latter on personal diversity preferences, specifically in employees’ selection into hypothetical workplaces. In a survey study with 364 European American respondents in three waves, participants complied with social consensus preferences for either more or less workplace diversity. The new preference was sufficiently internalized to be retained largely unaltered a week later. Simulations suggest a self-reinforcing effect, where accurate social consensus information may be sufficient to change preferences. Given that initial choices were polarized, perceived social consensus can vary highly between people in society, and influencing this perception may feed back into greater acceptance of minorities.

ARTICLE HISTORY
Received 1 February 2018
Accepted 19 October 2018

KEYWORDS
Social influence; segregation; workplaces; ingroup bias; status

Introduction
There is widespread segregation (by race, gender etc.) in society across several domains of our everyday lives, including segregation between workplaces (see, e.g., Åslund & Skans, 2010; Stainback & Tomaskovic-Devey, 2012), residential areas (e.g., Charles, 2003) and schools (e.g., Reardon & Owens, 2014; Saporito & Lareau, 1999), decreasing contact between people of different backgrounds and limiting the opportunity structure for large groups of people (e.g., Massey & Denton, 1993).

Segregation is not only a social problem, but it is also potentially a social dilemma. For example, it was illustrated in a famous theoretical model by Schelling (1969, 1971) that even if people were to prefer to live in mixed societies, society will still end up being residentially segregated if the preference for having at least a few group peers around you is stronger. This prediction has been confirmed through empirical tests of the model (e.g., Clark, 1991) and through simulation models based on empirical data on ethnoracial residential preferences (e.g., Clark & Fossett, 2008; Zhang, 2004) and has been shown to hold for different preference functions (van de Rijt, Siegel, & Macy, 2009) and for school segregation (e.g., Spaiser et al., 2018). Recent empirical work has also produced results consistent with this theoretical model for diversity preferences in a hypothetical scenario of choosing among potential workplaces: European American (the majority group) respondents favored...
workplaces with a considerable share of minority group employees, but workplaces where they would themselves be in the numerical minority were highly unpopular (Bursell & Jansson, 2018). Simulations also show that this could indeed produce high levels of segregation. Thus, levels of segregation are higher than what would be preferred even by privileged individuals.

One way to decrease segregation is if individual preferences would become even more diversity-prone. In the case of the aforementioned empirical study, this would mean that respondents would choose hypothetical workplaces with more outgroup members. From experiments in social psychology (e.g., Sechrist & Stangor, 2005), we know that prejudice can be reduced by social consensus information, that is, information about other people’s preferences (similar to the definition of Sechrist & Milford, 2007, where the same term refers to information about other people’s beliefs). Apart from homophily and prejudice, another reason to avoid vastly diverse workplaces could be status beliefs (Ridgeway, Boyle, Kuipers, & Robinson, 1998), and that these workplaces may be perceived as having lower status. Also in this case, social consensus information should be effective, if it delivers the message that diverse workplaces are more (or less) popular than expected. Given the observation that preferences are more diversity-prone than the macrolevel segregated outcome, it might suffice to receive truthful information about others’ preferences, to change what might be an unpopular norm (cf. Centola, Willer, & Macy, 2005; Thaler & Sunstein, 2008). With either underlying psychological mechanism, social influence could have a significant impact on diversity preferences, and ultimately on segregation outcomes.

We will here focus on segregation between workplaces, and more specifically preferences for potential colleagues. In a workplace where there is at least some teamwork, your colleagues are also your everyday interaction partners. In contrast, citizens in a large neighborhood rarely interact, and school choices are often made by someone else than the students themselves. Even if status beliefs and prejudice are important in all three domains of segregation, this suggests that any diversity preferences relating to workplace choices may better reflect personal preferences beyond status beliefs.

In a previous study, we showed that the stated preferences for diversity among American respondents would preserve or even lead to increased segregation, if realized (Bursell & Jansson, 2018). In this paper we move beyond this question and investigate the influence of social consensus information on preferences for diversity at the workplace. Given information on others’ preferences, will respondents update their own stated preferences, and how? To go beyond stated preferences and mere compliance, we also test whether any such updated preferences are sufficiently internalized to remain present some time after the social consensus information was given. Finally, what are the predicted long-term consequences for the social consensus of individuals being exposed to representative or unrepresentative social consensus information? Could accurate information also increase or decrease preferences for diversity?

**Theory and previous research**

Our study addresses the social problem of workplace segregation between ethnoracial groups, and draws upon theories of homophily and status construction to suggest that social influence may be used to achieve bias reduction.
Ethnoracial workplace segregation

American workplaces are segregated along racial lines, beyond random allocation (Stainback & Tomaskovic-Devey, 2012). This has many explanatory components: historical and contemporary discrimination, socioeconomic inequalities creating unequal life chances, and so on. However, over the course of a working life, employees are likely to have at least bounded influence over where they work. Even weak ethnoracial biases have the potential of generating almost complete segregation (e.g., Schelling, 1969, 1971). Hence, employees could also impact workplace segregation by their tenure at a workplace: by staying longer at workplaces where they work with similar others, and leaving workplaces where they constitute a minority. We expand upon the literature on segregation, first by studying the preferences among employees with respect to workplace segregation, and second, by testing the malleability of workplace preferences through social information (e.g., Salganik, Dodds, & Watts, 2006).

Homophily, status and bias reduction

Findings indicating that people prefer to surround themselves with ingroup members are generally explained by social identity theory. Since it is important for people to have a positive self-image, people strive to maintain a positive view of themselves and their ingroup (Ellemers, Spears, & Doosje, 2002; Hogg & Abrams, 1988). This may be achieved by perceiving the ingroup as positively distinct from other groups (Tajfel & Turner, 1979), which is easier in a homogenous social environment (Rubin & Hewstone, 1998).

But if homophily is important, why would people be susceptible to social influence? One explanation comes from status construction theory (c.f. Ridgeway, 1991; Ridgeway et al., 1998). It argues that status is another mechanism that influences people’s attitudes and preferences – the different statuses ascribed to social groups affect whom individuals want to associate with at workplaces. Cultural status beliefs about groups of people shape individuals’ social relations through three processes with consequences for inequality: status biases that shape implicit assumptions about who is ‘better’ in various ways; associational preference biases that shape whom people want to be associated with; and resistance reactions to status challenges (Ridgeway, 2014). When a society is stratified along ethnoracial lines, ethnoracial workplace composition becomes a marker of job status, a status that, according to status construction theory, cannot be reduced to questions of power or resources (Ridgeway, 2014). Because of cultural beliefs about the general competence of different social categories, a high proportion of individuals from the subordinate minorities signals a low workplace status. To the extent that signals of job status explain discrimination against diverse workplaces, social consensus information on such workplaces being popular should influence the decisions of potential employees to work there, by decreasing the status bias against them.

Social influence

People look to social norms to gain an accurate understanding of, and effectively respond to, social situations (Cialdini, 2003), with a desire to fit in and avoid deviating
from the norm. Crandall, Eshleman, and O’Brien (2002) found that the public expression of prejudice towards many groups correlated highly with social approval of that expression. However, social information is used not only to comply and fit in, but also has the effect of updating public attitudes, for example antiracist opinions (Blanchard, Crandall, Brigham, & Vaughn, 1994). While majority views exert more social pressure, targets are also more likely to assume that majorities are correct (Abrams & Hogg, 1990), so social consensus information on the most common preference should also update people’s information about the world.

Social influence often takes on subtler forms than outright peer pressure. Several experiments have demonstrated that stereotypes and prejudice can be altered, at least temporarily, purely by presenting subjects with the beliefs of others (for a review, see Sechrist & Stangor, 2005). These studies suggest that European American university students’ stereotypes of minorities are susceptible to social consensus information, also with a lasting effect (Stangor, Sechrist, & Jost, 2001), and that social information can impact European American students’ behavior towards African Americans (Sechrist & Milford, 2007; Sechrist & Stangor, 2001). Beliefs about norms and other people’s behavior can also influence honesty (Rauhut, 2013) and energy conservation (Schultz, Nolan, Cialdini, Goldstein, & Griskevicius, 2007), in the latter case through presenting average neighborhood usage.

One of our aims is to assess whether the effects from social consensus information apply also to preferences (i.e., the ordering of alternatives, based on desires and satisfaction), and to more specific and personally relevant preferences – those for diversity at the subject’s workplace. This situation differs from previous measures within the social influence literature, on stereotypes, acquiring information and updating beliefs, as the subjects need to consider the situation of them interacting with minorities in their everyday life. Workplace affiliation is for many people part of personal identity and may, as opposed to rating average characteristics of outgroups, create a more emotionally grounded response, being relevant to the personal sphere.

**Internalization and long-term effects**

In terms of the varieties of social influence identified by Kelman (1958), *compliance*, when people conform while keeping dissenting opinions private, will potentially change people’s perception of social consensus, but to change preferences and actual behavior, some *internalization*, when people accept and agree privately, is needed.

We hypothesize that people’s preferences for ethnoracial diversity in the workplace are liable to be updated in either direction towards common responses by other respondents. We test whether such an effect is the result of pure compliance when being given the information, or there is sufficient internalization for it to remain one week later. We also investigate potential long-term effects, and hypothesize that presenting accurate feedback also has the potential of making respondents more diversity-prone, to the extent that they previously expected preferences to be less so. Thus, we direct our attention towards how social psychological mechanisms at the individual level may impact social change (c.f. Abrams & Vasiljevic, 2014), also looking into the feedback between attitude polarization (i.e., when people cluster w.r.t. attitudes and the distance between the attitudes of the clusters increases) and social influence.
Methods

The study was carried out in three waves of surveys presented to the same respondents with approximately one week in between. The participants were not told that there would be follow-up studies. Respondents, who were paid $0.4 for each completed survey, were recruited at the Amazon Mechanical Turk.

In the first wave, the respondents were asked to reflect upon what their ideal workplace (c.f. Brown-Iannuzzi, Payne, & Trawalter, 2012) would look like in terms of ethnoracial and gender composition (i.e., a measure of preferences) in a number of tasks.¹ In the task of interest here, we presented a list of five hypothetical companies, each with eight employees. The only information given about the companies was that X had 8 European American and 0 minority group employees, Y had 6 vs. 2, Z had 4 vs. 4, W had 2 vs. 6, and V had 0 vs. 8. The task given to the respondents was to rank the companies in descending order according to their preference for working there.

About a week later, the respondents were reinvited to perform (only) the rankings task, which this time included the manipulation of the experiment. Along with the instructions for the task, the respondents were also presented with results from the previous wave, followed by one randomly selected ranking out of three. The three conditions were the rankings 02468 (i.e., companies X, Y, Z, W, V, in that order, which represents minority avoidance, or low diversity), 42068 (even mix with avoidance of self being in the minority, or medium diversity) and 46280 (even mix with a secondary preference for minorities, or high diversity). The respondents were informed that the previous rankings information was not necessarily representative.² There was also a control treatment, in which no social information was presented, but the task from the previous wave was just repeated.

Finally, about a week later again, those who completed the second survey were reinvited to perform the rankings task once again. As in the first round, there was no social information given. This third task allowed us to test whether any effects from social information were lingering a week after the information had been given, or whether the participants had reverted to their original preferences, and thus, importantly, whether choices were updated only in connection to the stimulus, or they remained.

The first wave included variables that are not analyzed here, but in a different paper (Bursell & Jansson, 2018). For the task of interest here, the ranking of companies are the only variables in the three waves. All respondents in a wave were reinvited to participate in the next wave, but only those (US born European Americans) completing all waves are included in the analysis (apart from the drop-out analysis, which includes all respondents). Among those, there are no missing values.

Using the measured changes in preferences between the first and third occasion, we conducted simulations to investigate the long-term effects from the suggested dynamics of these experiments, were they to be conducted an indefinite number of rounds, assuming that the effects remain the same.

Measuring rankings

Even though we will also look at rankings directly, there are strong statistical dependencies in rankings data. Ranking one item limits the choices for the remaining ones, and the
last item is completely dependent on the rankings of the first ones. Partly to address this issue, we will also look at the scale measure of counting the number of *inversions* in the permutation (elements ordered in a sequence) that the ranking represents.

In mathematics, an ordered pair of numbers in the ranking is an inversion if the greater number comes before the smaller number. The inversion number is thus the total number of pairs that are not arranged in an ascending order. The inversion count gives us a measure of minority preference, ranging from the ranking 02468 with zero inversions to the ranking 86420 with ten (all pairs are inversions). The ranking 42068 has three inversions: 42, 40 and 20, and similarly 46280 has six. Thus, with this measure, 0 represents maximizing the number of European American colleagues, 10 maximizing the number of minority group colleagues, and 5 having no preference between these. The number of inversions answers the question ‘in how many of all pairs of companies would you choose the more diverse over the less diverse company?’, which is a scale variable that also has a *meaningful* interpretation as such (Marcus-Roberts & Roberts, 1987), since the comparison is in turn over the scale variable number of minority employees, with equidistant answer options.

It should be noted that even Likert scales can arguably be analyzed as scale variables, with parametric tests (see e.g., Norman, 2010). Contrasting to Likert scales, the intervals between each number of inversions are here equal by definition, and the meaning of increasing the inversion count by one is clearly defined: it means ranking *one* (more) company lower than *one* less diverse company (conversely, moving a company *one* step in ranking results in *one* more or fewer inversion, depending on whether the company that became higher ranked is more or less diverse). The intervals between companies are in turn equal, where the most similar more diverse company has two more minority group employees. In can also be noted that our social information manipulations are equidistant in terms of the number of inversions.

**Results**

There were 1,068 respondents completing the task correctly (out of 1,100) on the first occasion, out of which 702 completed it on the second, and 510 also on the third occasion. We included only European Americans since the workplace compositions to consider in the survey relate to this group versus members of any minority group. In the first round, 762 respondents were born in the US and identified as European Americans. A total of 364 of these completed all three rounds, and our analyses will pertain to these.

Respondents dropping out before completing all three surveys are thus excluded, and the personal background information of these respondents is similar to that of those completing the study. The proportion of female respondents is 41% (both in the first and the third round). The mean age is 35 years (33 in the first round) and the median 32 years (30 in the first round). A total of 97% have at least one year of work experience, with a median of 12 years. A logistic regression model revealed no significant effect from number of inversions in the first wave on participating in the second wave (OR = 1.01, $\chi^2(1) = 0.18$, $p = .67$), and no effect from inversions in the first wave, the three information conditions, and interactions between initial inversions and condition on participating in the third wave (all ORs are in the interval (0.74,1.14), $\chi^2(7) = 1.8$, $p = .97$), so we find no attrition bias.
Descriptives

Figure 1 presents average rankings in the first round among native-born European American respondents. The average number of minority group members in companies with respect to ranking goes from low to high, where the company in the respondents’ first and second rank have fewer minority group employees than in the third rank, and so on, revealing a preference for low diversity.

Meanwhile, there is a clear preference for medium and low diverse companies to those void of minorities. Figure 2 shows the proportional frequencies of each choice, where 2 and 4 are chosen over 0 in the majority of cases in the left panel. There is an almost unanimous agreement that companies with only minority group employees are the least desirable ones. The aggregated preferred ranking from pairwise comparisons in the left panel, or, likewise, cumulating the numbers in the right panel, is 42068. The average number of inversions in the rankings of the respondents is 2.80, which is close to 3, the number of inversions in the aggregated preferred ranking.

Figure 2 also reveals some polarization. The company with two minority employees is ranked second in the majority of cases, while the one ranked first has either more or less diversity. Table 1 depicts the six most commonly chosen rankings, and provides further evidence of polarization. While the average ranking is one that optimizes

![Figure 1](image-url)  
*Figure 1.* The average number of minority group employees in the respondents’ first ranked company, second ranked, etc., with regular standard error of the mean bars.*
diversity without the respondent ending up as being in the ethnoracial minority, the most common ranking is the one that avoids minorities, followed by one that stays as close to a fifty-fifty distribution as possible (with a preference for European Americans when forced to choose).

The interpretation of a ranking is straightforward as long as it is consistent. There were 26 respondents (7%) providing 21 different inconsistent rankings (such as 04268), so most respondents do provide easily interpretable rankings, and there is no inconsistent ranking that the remaining respondents agree on.

Based on these results, we can say something more on the representativity of the social information presented in the second survey round. The ranking 42068 is also the average ranking. The ranking 02468, with three fewer inversions, is the most common one. Finally, 46280, with three more inversions than the average, is at the extreme of how much diversity respondents prefer; only seven percent have chosen rankings with more inversions. At the same time, it shares the property with the more common 42608
that companies as close to a fifty-fifty distribution as possible are preferred. We thus argue that neither of the rankings presented in the social information is incredible.\textsuperscript{6}

**Experiment**

An analysis of variance over the four conditions and three waves showed no significant effect of condition, $F(3,360) = 1.66$, $MSE = 13.9$, $p = .18$, nor of time, $F(2,720) = 0.77$, $MSE = 1.04$, $p = .46$. However, there is a significant interaction effect, $F(6,720) = 6.09$, $MSE = 1.04$, $p < .001$. The aggregated results from each social information condition and wave are presented in the left panel of Figure 3. The error bars are 95% within-subject confidence intervals\textsuperscript{7} (Cousineau, 2005; Loftus & Masson, 1994; Morey, 2008). The figure suggests a positive effect of high diversity, 46280 ($M = 0.73$, $SD = 1.69$, difference between first and second round), $t(90) = 4.09$, $p < .001$, and a negative one of low diversity, 02468 ($M = -0.38$, $SD = 1.51$), $t(87) = -2.33$, $p = .011$, one-sided (since we have a directed alternative hypothesis), on the number of inversions, while there is no clear effect from medium diversity, 42068 ($M = 0.11$, $SD = 1.61$), $t(94) = .64$, and the control condition ($M = -0.26$, $SD = 1.34$), $t(89) = -1.81$. The significant effects remain significant with Holm–Bonferroni adjusted $\alpha$ ($\geq .0125$) levels (Holm, 1979).

However, the respondents for each of these conditions include those who had initially more as well as those who had fewer or an equal number of inversions compared to the information given, that is, for some respondents, the social consensus information was more diversity-prone than was the respondent, while for others it was less.

Separating these into different conditions and retaining those with enough respondents, we retain five conditions: low−, respondents who received the social consensus ranking 02468 and had more than 0 inversions in their own initial ranking, thus receiving negative social information, in terms of diversity preference, compared to their own preferences; medium−, with respondents with more than 3 inversions;
medium+, with respondents with fewer than 3 inversions; high+, with respondents with fewer than 6 inversions; and a control condition with no social information.

The right panel in Figure 3 shows the average individual change in the number of inversions for each of these conditions between the first and the second, and the first and the third round, together with 95% confidence intervals for the change from the first round. The figure shows that for all conditions, except the control condition, the number of inversions changes significantly in the predicted direction, and, most importantly, the new preferences displayed in the second round are retained in the third round (differences between the first and third round: low−: $M = -0.42$, $SD = 1.54$, $t(70) = -2.32$, $p = .012$; medium−: $M = -0.63$, $SD = 1.73$, $t(42) = -2.38$, $p = .011$; medium+: $M = 0.93$, $SD = 1.60$, $t(41) = 3.76$, $p < .001$; high+: $M = 0.82$, $SD = 1.75$, $t(77) = 4.14$, $p < .001$, all one-sided, since the alternative hypotheses are directed, and significant against Holm–Bonferroni adjusted $\alpha \geq .0125$), largely unchanged from the second round (differences between the second and third round: low−: 0.10, $t(70) = 0.67$; medium−: −0.07, $t(42) = -0.44$; medium+: 0.07, $t(41) = 0.30$; high+: −0.04, $t(77) = -0.38$; control: 0.10, $t(89) = 1.07$), without showing any regression towards the mean. Thus, the social consensus information presented to the respondents does alter their stated preferences in line with the information, and the new preferences remain largely unaltered a week later. The conditions with information that is more positive towards diversity than the expressed preferences of the respondent show a greater measured effect than those with negative information.

Although not significant, the control condition also had a negative estimate of change. In order to compare to the control condition and control for initial number of inversions, a linear regression model of the change, between the first and the third round, dependent on initial preferences, experimental condition and an interaction of these is presented in Table 2 ($F(7,356) = 11.7$, $MSE = 2.20$, $p < .001$, $R^2 = .19$). Neither intercept nor slope for the low condition is significantly different from those in the control treatment, suggesting that the added effect from negative information is small. Both the other conditions have significantly larger intercepts, and, as hypothesized from the fact that medium constitutes either positive or negative information dependent on initial preferences, positive change decreases significantly with the subject’s initial number of inversions.

An expanded linear regression model including sex, age and having a university degree, including interaction terms with first wave inversions and condition, did not

|               | Estimate | Std. error | p      |
|---------------|----------|------------|--------|
| Inversions    | .251     | .251       | .317   |
| Control       | -.158    | .076       | .038*  |
| Low (02468)   | .085     | .373       | .819   |
| Medium (42068)| 1.123    | .354       | .002** |
| High (46280)  | .861     | .331       | .010** |
| Inversions * control | -.048 | .107       | .650   |
| Inversions * low | -.244  | .099       | .014*  |
| Inversions * medium | -.033 | .097       | .731   |
| Inversions * high | -.033  | .097       | .731   |
produce a significantly reduced residual sum of squares \( F(24,329) = 1.25, \text{MSE} = 0.20, p = 0.20 \). We conclude that these variables are unlikely to be important mediators of changing preferences.

As a note on the inversions measure, a positive (negative) change may result from some diverse companies being more (less) preferred than previously, in which case the change is easy to interpret qualitatively, or some diverse companies may have been ranked higher, while others have been ranked lower. Such inconsistent changes occurred in 8% of all cases, so most changes consisted of only increasing or only decreasing diverse companies in rank.

**Long-term dynamics and polarization**

We ran simulations to investigate what would be the long-term consequences if we were to continue this process over even more rounds, with accurate social consensus information (see the Appendix). In sum, assuming that people’s reaction to social consensus information can be generalized by our regression model, we make three observations: (1) people’s opinions converge, giving homogeneity in preferences, (2) the effect is generally a positive movement towards diversity preferences, but (3) with information dispersal in clusters, the population can easily polarize into clusters with more or less diversity-prone preferences.

**Conclusions**

In a repeated measures experiment, we found, in line with social identity theory, that majority group members prefer workplaces with mostly majority group employees, but with some minority representation, and that these preferences are susceptible to social consensus information, at least for more diversity. The new stated preferences were retained in the last round, a week after the social consensus information was given, whether it was positive or negative.

Drawing on these results, our study speaks to the literature on prejudice reduction by providing additional evidence that social influence can be used to create durable effects concerning a related topic: ethnoracial preferences. In line with, and expanding upon, previous findings of week- to month-long effects for ethnoracial beliefs and prejudice (Stangor et al., 2001; Zitek & Hebl, 2006), social consensus information can change people’s preferences for having minority group members around themselves at the workplace, also beyond the intervention occasion.

From the second wave, where the respondents updated their preferences when being provided information on previous results, we could only infer compliance with what might be perceived as social norms. However, the fact that new preferences remained largely unaltered a week later is consistent with the idea that some internalization is going on. Workers at the Amazon Mechanical Turk generally participate in many tasks each day (some counting by the hundreds), and we find it likely that they would not remember information from a specific task a week earlier on a full ranking of companies, and just reiterate their previous responses. An alternative explanation, if the workers do actually remember their responses, is that rankings are preserved between the second and third round due to internal consistency, a desire to avoid cognitive dissonance and maintain values over time (Festinger, 1957). However, such an explanation is at least not sufficient, since there is a significant change in
stated preferences between the first and second round. It may be that the respondents remember a general tendency towards including more or fewer minority group employees, and that more is required for updating preferences internally, but remembering and acting upon this in absence of the stimulus, is at least a first step towards internalization.

The results are also consistent with status construction theory. If companies are selected against due to low expected status, then social consensus information updates expected status. Ranking of companies may thus reflect beliefs of status, and, ultimately, beliefs about other people’s beliefs of status (De Kwaadsteniet, Homan, Van Dijk, & Van Beest, 2012; Jansson & Eriksson, 2015). Ridgeway, Backor, Li, Tinkler, and Erickson (2009) have shown empirically that experiences of social differences are very easily transformed into status beliefs, and that individuals (at least men) tend to act upon these beliefs. The social information presented in our study may be perceived as a status queue, a hint on what is a desirable workplace. If this is the case, then the mechanism explaining the respondent’s preference change has its roots in an adjustment made to maximize future status increase rather than in genuine normative attitude change. This mechanism is thus distinct from the one explaining the changes observed by Stangor et al. (2001) and Zitek and Hebl (2006), which concerned adjustment to attitudes and stereotypes concerning equality. While we cannot, based on our data, identify which of these mechanisms is the primary driver of our results, future research could possibly disentangle them through new experiments where status preference is separated from preferences sensitive to social desirability.

In line also with the argument that workplace segregation is a social dilemma with segregation beyond individual preferences, which may in turn create an expectation among individuals that the social consensus preferences are for less diversity, accurate information seemed sufficient to increase individual diversity preferences. This may of course not be sufficient for desegregation. While the segregation outcomes from avoiding being in the vast minority (Schelling, 1971) can produce an underestimation of diversity preferences in the population, preferences for diversity may need to be increased considerably for the same process not to produce a segregated outcome (see also van de Rijt et al., 2009).

Our findings show that workplace preferences may evolve through feedback loops between individual preferences and perceptions of aggregated preferences. If the perceptions are accurate, then we would expect preferences to reach an equilibrium. In our simulations (in the Appendix), this equilibrium involved a higher acceptance of minorities than initially. However, initial private opinions were polarized (Table 1), and, depending on your social network, you may be exposed to extreme, rather than average, opinions and move in the corresponding direction. In the simulations, we saw that, even in well-mixed populations that are not initially clustered based on preferences, given either extreme information conditions or small populations, we would see polarization of opinion where the clusters of connected people reach different social consensus equilibria.

The demonstrated effects of social consensus information may have policy potential for counteracting different types of segregation and other social problems where ethnoracial preferences or status aspirations contribute to a negative aggregated outcome. Both diversity-averse and diversity-prone preferences are prevalent in the population. By exposing the latter as social consensus, as in our simulations, the preferences of more people might move in that direction, feeding back into moving even more people in the same direction, as social consensus information, derived from previous choices, updates.
Notes

1. The other two tasks, which were given before the present one, were to rate their general feeling towards working with someone from each of the groups African American/Black/Caribbean/African, Asian/Pacific Islander, European American, Hispanic/Latino, Middle Eastern/North African and Native American/Other, divided into men and women, giving a total of twelve groups; and to compose an ideal workplace in terms of proportional shares of the groups European American and minority group men and women. The results from these tasks (which were given only in the first wave) are presented in Bursell & Jansson (2018), where the survey in the first wave can also be found. Completing the other two tasks before the rankings task defined the various minority groups, prompted respondents to think about their relation to them, also in relation to which would be their first choice in terms of ideal colleagues, before ranking preferable outcomes when the first choice is not available. The second and third wave included only the rankings task.

2. ‘We have run this survey previously at the Amazon Mechanical Turk. In a sample of 100 participants, the most common ranking was the following:

3. For example, 123 has 0 inversions, 132 has 1 (32), 312 has 2 (31 and 32) and 321 has 3 (32, 31 and 21).

4. The question can be equivalently formulated as ‘How many minority employees would you choose over majority employees’, where each inversion would thus count double and provide a measure between 0 and 20.

5. We define a consistent ranking as one where a company with \(n\) minority employees is ranked such that \(n\) does not differ by more than two from the number of minority employees at all the more preferred companies. E.g., 46208 is consistent, but 46028 is not, since 0 differs by more than two from both 4 and 6. A consistent ranking represents a primary preference for a certain mix and secondary preferences away from that mix without jumps, i.e., either increasing or decreasing diversity gradually, staying as close as possible to the primary preference.

6. In the free comment section no one expressed any suspicion (though some were surprised by the information). The first two rankings can both be argued to be accurate depictions of peer preferences. The last ranking does not represent a random sample, but is the most common one for, say, the group of 104 minority group respondents (not included in the analysis) with at least some college education and a foreign-born father.

7. The confidence intervals are between subjects, and thus for comparison between waves, not conditions.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by the FP7 Ideas: European Research Council [324233]; Swedish Research Council for Health, Working Life and Welfare [2016-00173]; Knut and Alice Wallenberg Foundation [2015.0005]; Riksbankens Jubileumsfond [M12-0301:1]; and the Swedish Research Council [340-2013-5460,421-2012-1027,445-2013-7681]

ORCID

Fredrik Jansson http://orcid.org/0000-0001-8357-0276
Moa Bursell http://orcid.org/0000-0002-8289-4401
References

Abrams, D., & Hogg, M. A. (1990). Social identification, self-categorization and social influence. European Review of Social Psychology, 1(1), 195–228.

Abrams, D., & Vasiljevic, M. (2014). How does macroeconomic change affect social identity (and vice versa?): Insights from the European context. Analyses of Social Issues and Public Policy, 14(1), 311–338.

Åslund, O., & Skans, O. N. (2010). Will I see you at work? Ethnic workplace segregation in Sweden, 1985-2002. Industrial and Labor Relations Review, 63(3), 471–493.

Blanchard, F. A., Crandall, C. S., Brigham, J. C., & Vaughn, L. A. (1994). Condemning and condoning racism: A social context approach to interracial settings. Journal of Applied Psychology, 79(6), 993–997.

Brown-Iannuzzi, J. L., Payne, B. K., & Trawalter, S. (2012). Narrow imaginations: How imagining ideal employees can increase racial bias. Group Processes & Intergroup Relations, 16(6), 661–670.

Bursell, M., & Jansson, F. (2018). Diversity preferences among employees and ethnoracial workplace segregation. Social Science Research, 74, 62–76.

Centola, D., Willer, R., & Macy, M. (2005). The emperor’s dilemma: A computational model of self-enforcing norms. American Journal of Sociology, 110(4), 1009–1040.

Charles, C. Z. (2003). The dynamics of racial residential segregation. Annual Review of Sociology, 29(1), 167–207.

Cialdini, R. B. (2003). Influence: Science and practice (4th ed.). Boston, MA: Allyn & Bacon.

Clark, W. A. V. (1991). Residential preferences and neighborhood racial segregation: A test of the Schelling segregation model. Demography, 28(1), 1–19.

Clark, W. A. V., & Fossett, M. (2008). Understanding the social context of the Schelling segregation model. Proceedings of the National Academy of Sciences of the United States of America, 105(11), 4109–4114.

Cousineau, D. (2005). Confidence intervals in within-subject designs: A simpler solution to Loftus and Masson’s method. Tutorials in Quantitative Methods for Psychology, 1(1), 42–45.

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.

De Kwaadsteniet, E. W., Homan, A. C., Van Dijk, E., & Van Beest, I. (2012). Social information as a cue for tacit coordination. Group Processes & Intergroup Relations, 15(2), 257–271.

Ellemers, N., Spears, R., & Doosje, B. (2002). Self and social identity. Annual Review of Psychology, 53(1), 161–186.

Festinger, L. (1957). A theory of cognitive dissonance. Redwood City, CA: Stanford University Press.

Hogg, M. A., & Abrams, D. (1988). Social identifications: A social psychology of intergroup relations and group processes. London: Routledge.

Holm, S. (1979). A simple sequentially rejective multiple test procedure. Scandinavian Journal of Statistics, 6(6), 65–70.

Jansson, F., & Eriksson, K. (2015). Cooperation and shared beliefs about trust in the assurance game. PLoS ONE, 10(12), e0144191.

Kelman, H. C. (1958). Compliance, identification, and internalization three processes of attitude change. Journal of Conflict Resolution, 2(1), 51–60.

Loftus, G. R., & Masson, M. E. J. (1994). Using confidence intervals in within-subject designs. Psychonomic Bulletin & Review, 1(4), 476–490.

Marcus-Roberts, H. M., & Roberts, F. S. (1987). Meaningless Statistics. Journal of Educational Statistics, 12(4), 383–394.

Massey, D. S., & Denton, N. A. (1993). American apartheid: Segregation and the making of the underclass. Cambridge, MA: Harvard University Press.

Morey, R. D. (2008). Confidence intervals from normalized data: A correction to Cousineau (2005). Tutorials in Quantitative Methods for Psychology, 4(2), 61–64.

Norman, G. (2010). Likert scales, levels of measurement and the “laws” of statistics. Advances in Health Sciences Education, 15(5), 625–632.
Rauhut, H. (2013). Beliefs about lying and spreading of dishonesty: Undetected lies and their constructive and destructive social dynamics in dice experiments. *PLoS ONE*, 8(11), e77878.

Reardon, S. F., & Owens, A. (2014). 60 years after brown: Trends and consequences of school segregation. *Annual Review of Sociology*, 40(1), 199–218.

Ridgeway, C. L. (1991). The social construction of status value: Gender and other nominal characteristics. *Social Forces*, 70(2), 367–386.

Ridgeway, C. L. (2014). Why status matters for inequality. *American Sociological Review*, 79(1), 1–16.

Ridgeway, C. L., Backor, K., Li, Y. E., Tinkler, J. E., & Erickson, K. G. (2009). How easily does a social difference become a status distinction? Gender matters. *American Sociological Review*, 74(1), 44–62.

Ridgeway, C. L., Boyle, E. H., Kuipers, K. J., & Robinson, D. T. (1998). How do status beliefs develop? The role of resources and interactional experience. *American Sociological Review*, 63(1), 1–16.

Rubin, M., & Hewstone, M. (1998). Social identity theory’s self-esteem hypothesis: A review and some suggestions for clarification. *Personality and Social Psychology Review*, 2(1), 40–62.

Salganik, M. J., Dodds, P. S., & Watts, D. J. (2006). Experimental study of inequality and unpredictability in an artificial cultural market. *Science*, 311(5762), 854–856.

Saporito, S., & Lareau, A. (1999). School selection as a process: The multiple dimensions of race in framing educational choice. *Social Problems*, 46(3), 418–439.

Schelling, T. C. (1969). Models of segregation. *The American Economic Review*, 59(2), 488–493.

Schelling, T. C. (1971). Dynamic models of segregation. *The Journal of Mathematical Sociology*, 1(2), 143–186.

Schultz, P. W., Nolan, J. M., Cialdini, R. B., Goldstein, N. J., & Griskevicius, V. (2007). Destructive, and reconstructive power of social norms. *Psychological Science*, 18(5), 429–434.

Sechrist, G. B., & Milford, L. R. (2007). The influence of social consensus information on intergroup helping behavior. *Basic and Applied Social Psychology*, 29(4), 365–374.

Sechrist, G. B., & Stangor, C. (2001). Perceived consensus influences intergroup behavior and stereotype accessibility. *Journal of Personality and Social Psychology*, 80(4), 645–654.

Sechrist, G. B., & Stangor, C. (2005). Prejudice as social norms. In C. S. Crandall & M. Schaller (Eds.), *Social Psychology of Prejudice: Historical and Contemporary Issues*, 167–187. Lawrence, KS: Lewinian Press.

Spaiser, V., Hedström, P., Ranganathan, S., Jansson, K., Nordvik, M. K., & Sumpter, D. J. T. (2018). Identifying complex dynamics in social systems. *Sociological Methods & Research*, 47(2), 103–135.

Stainback, K., & Tomaskovic-Devey, D. (2012). Documenting desegregation: Racial and gender segregation in private sector employment since the civil rights act. New York, NY: Russell Sage.

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.

Tajfel, H., & Turner, J. C. (1979). An integrative theory of intergroup conflict. In W. G. Austin & S. Worchel (Eds.), *The social psychology of intergroup relations* (pp. 33–47). Monterey, CA: Brooks-Cole.

Thaler, R. H., & Sunstein, C. R. (2008). *Nudge: Improving decisions about health, wealth, and happiness*. New Haven, CT: Yale University Press.

van de Rijt, A., Siegel, D., & Macy, M. (2009). Neighborhood chance and neighborhood change: A comment on Bruch and Mare. *American Journal of Sociology*, 114(4), 1166–1180.

Zhang, J. (2004). Residential segregation in an all-integrationist world. *Journal of Economic Behavior & Organization*, 54, 533–550.

Zitek, E. M., & Hebl, M. R. (2006). The role of social norm clarity in the influenced expression of prejudice over time. *Journal of Experimental Social Psychology*, 43(6), 867–876.
Appendix

The direct analyses of the data have given us a direction of change based on social influence: respondents become more similar to social consensus. What would be the long-term consequences if we were to continue this process over even more rounds, and provide accurate social consensus information? Would preferences converge, and are there one or multiple equilibria? And would accurate social consensus information increase or decrease preferences for diversity?

To incorporate our data as initial conditions, we designed and implemented a simple simulation model. The model assumes a population of \( N \) agents with diversity preferences represented by a number of inversions \( x_I \) sampled from the distribution given by our data (i.e., we model inversions, not rankings). In each round, a random agent is exposed to the social consensus information condition from the experiment (i.e., low with 0, medium with 3 or high with 6 inversions) that is closest to the mean number of inversions of the agents’ preferences. This represents accurate social information with low resolution. The agent updates its preference by an amount drawn from a normal distribution with mean and standard deviation given by the regression model in Table 2 in the paper.

The results for 3.10 simulation runs are presented in Figure A1. In the left and middle panels, the population size is set to \( N = 100 \), in the right panel \( N = 25 \). In the middle panel, we have excluded the moderate social consensus information condition, with 3 inversions, providing only extreme information.

In the first set of runs, social consensus quickly converges to the middle equilibrium (which is slightly larger than the initial social consensus). In the second set of runs, including only the two extreme information conditions, we get polarization, where some populations end up in the higher equilibrium, and the remaining ones in the lower. The higher equilibrium appears to be stable, while the lower is unstable: some populations move above the threshold, and are quickly absorbed by the high equilibrium. Recall that diversity-prone information leads to higher social influence than diversity-averse information. Finally, in the third set of runs, with small populations, again we get polarization between runs, but with the middle and high equilibria, and some movement between both. In sum, assuming that people’s reaction to social consensus information can be generalized by our regression model, we make three observations: (1) people’s opinions converge, giving homogeneity in preferences, (2) the effect is generally a positive movement towards diversity preferences, but (3) with information dispersal in clusters, the population can easily polarize into clusters with more or less diversity-prone preferences.

In comparison, and being even more on the speculative side, what would happen with continuous social consensus information, not limited to three conditions? Assuming that we can inter- and extrapolate from the conditions with 0, 3 and 6 inversions, and represent these by

![Figure A1](image-url). Number of inversions over time in three sets of ten simulations each. In the left and right panel, all three social information conditions are available, with a small population in the right panel; in the middle panel, only the two extreme conditions are available.
a scale variable $x_C$ (instead of three dummy variables), we get the following regression model
(recall that $x_I$ is the initial inversion preference for the agent):

$$y = 0.621 - 0.295x_I + 0.107x_C + 0.00794x_Ix_C + \epsilon$$

If we solve this for $y = 0$ and $x_I = x_C = x^*$, then we get $x^* \approx 4.0$, that is, the population will
converge to a social consensus preference of four inversions, which is greater than the initial
mean of 2.8 in our data.