Personality and Social
Psychology Bulletin

Many mickles make a muckle: Evidence that gender stereotypes re-emerge spontaneously via cultural evolution

Journal:	*Personality and Social Psychology Bulletin*
Manuscript ID	PSPB-22-370.R3
Manuscript Type:	Empirical Research Paper
Keywords:	Stereotypes, Gender Stereotyping, Culture and Cognition, Social Cognition, Person Perception

Abstract:	We explore whether societal gender stereotypes re-emerge as social information is repeatedly passed from person to person. We examined whether peoples' memories for personality attributes associated with female and male social targets became increasingly consistent with societal gender stereotypes as information passed down social transmission chains. After passing through the memories of just four generations of participants, our initially gender-balanced micro-societies became rife with traditional gender stereotypes. While we found some evidence of the re-emergence of gender stereotypes in Expt. 1, we found the effects were stronger when targets appeared in a feminine stereotyped occupational context (Expt. 2) and in a masculine stereotyped occupational context (Expt. 3); conversely, the re-emergence of gender stereotypes was attenuated when targets appeared in a single gender context (Expt. 4). The current findings demonstrate that gender schematic memory bias, if widely shared, might cause gender stereotypes to be maintained through cultural evolution.

Many mickles make a muckle:

Evidence that gender stereotypes re-emerge spontaneously via cultural evolution

Carolyn J. Dallimore[1] Kenny Smith[2] Jacqui Hutchison[1]

Gillian Slessor[1] Douglas Martin[1]

[1]University of Aberdeen; [2]University of Edinburgh;

Word Count: 11728

Keywords: gender stereotypes; stereotype formation; cultural evolution; social cognition; person perception;

Author Note: None of the studies reported in this manuscript were pre-registered. The experiments we report were designed as a sub-project of a grant application (ES/N019121/1), awarded in 2016, before it was commonplace for research designs to be pre-registered; we state this explicitly in the manuscript.

Address Correspondence To:

Doug Martin

School of Psychology

Kings College

University of Aberdeen

Aberdeen, UK

AB24 3UB

+44 1224 273647

Email: doug.martin@abdn.ac.uk

Abstract

We explore whether societal gender stereotypes re-emerge as social information is repeatedly passed from person to person. We examined whether peoples' memories for personality attributes associated with female and male social targets became increasingly consistent with societal gender stereotypes as information passed down social transmission chains. After passing through the memories of just four generations of participants, our initially gender-balanced micro-societies became rife with traditional gender stereotypes. While we found some evidence of the re-emergence of gender stereotypes in Expt. 1, we found the effects were stronger when targets appeared in a feminine stereotyped occupational context (Expt. 2) and in a masculine stereotyped occupational context (Expt. 3); conversely, the re-emergence of gender stereotypes was attenuated when targets appeared in a single gender context (Expt. 4). The current findings demonstrate that gender schematic memory bias, if widely shared, might cause gender stereotypes to be maintained through cultural evolution.

Stereotypes; Gender Stereotyping; Culture and Cognition; Social Cognition; Person Perception;

"There is an old Scotch (sic) *adage…which none in the whole catalogue of them is more true or more worthy of being held in remembrance, viz,* "that many mickles make a muckle", *indicating that however trifling a thing may be in itself when it stands alone, when they come to be multiplied, they mount high…".* George Washington in a letter to James Germain, 1794 (as cited in *Writings of George Washington from the Original Manuscript*, 1931, pp.390)

Introduction

The aim of the current research was to establish whether societal gender stereotypes persist because stereotype-consistent memory bias incrementally shapes how social information evolves as it is repeatedly passed from person to person. It is striking that the content of gender stereotypes has remained relatively stable over time despite huge societal change, apparently-pervasive changes in people's attitudes, and extensive interventions to challenge stereotype content (Donnelly & Twenge, 2017; Eagly et al., 2020; Haines et al., 2016; Lueptow & Garovich-Szabo, 2001). Women are still stereotypically associated with *femininity, communion,* and *warmth*, whilst men are still stereotypically associated with *masculinity, agency,* and *competence* (Abele et al., 2020). When people endorse these gender stereotypes it can lead to overt prejudice and discrimination (e.g., Bem, 1981; Cundiff & Vescio, 2016; Devine, 1989); however, even if people do not endorse stereotypes, knowledge of their content can still lead to bias in thoughts and behaviour (e.g., Begeny et al., 2020; Bem, 1981; Greenwald & Banaji, 1995). While there is abundant evidence that stereotypes represent a fundamental source of social and cognitive bias, there remains a substantial gap in understanding how and why stereotype content remains so impervious to societal change (Ellemers, 2018). Addressing this gap, the current research investigated whether stereotype-consistent memory bias shapes the evolution of socially transmitted information, resulting in the spontaneous re-emergence of societal gender stereotypes that were initially absent from the social environment.

The central thesis of the current research is that even subtle cognitive bias can accumulate to exert substantial cumulative influence. The cumulative effects of small biases are cornerstones of many scientific disciplines; for example, evolutionary biology (e.g., where small selective differences drive the evolution of species), geology (e.g., where tiny incremental changes shape the Earth's topography), and climatology (e.g., where the decisions of individual humans collectively impact the global climate), provide notable examples of biases that are barely perceptible at the micro-level but that can have fundamental influence at the macro-level. As a discipline, psychology has been slower to embrace the potential macro-level societal effects that can occur as the cumulative consequence of micro-level biases in individual people (although see Bartlett, 1932; Mesoudi, 2011). This seems like a significant oversight. Human society is a consequence of a continuous cycle of cultural evolution[1] – people are exposed to information from their social environment, they represent this information cognitively, and they can communicate this information to other people, thereby creating a novel social environment from which others can learn (see Figure 1, left panel). We suggest every stage of this cycle is subject to bias, and when these biases are widely shared, they can exert directional pressure on the way information evolves – in the current context, bias towards stereotype-consistent cognitive representation exerts directional pressure that results in the re-emergence of gender stereotypes that were initially absent from the social environment.

[1] The term "cultural evolution" is broadly defined in the literature and has been applied to document how a diverse array of phenomena change over time, including but not limited to information, knowledge, behaviour, attitudes, and beliefs (see Mesoudi, 2011).

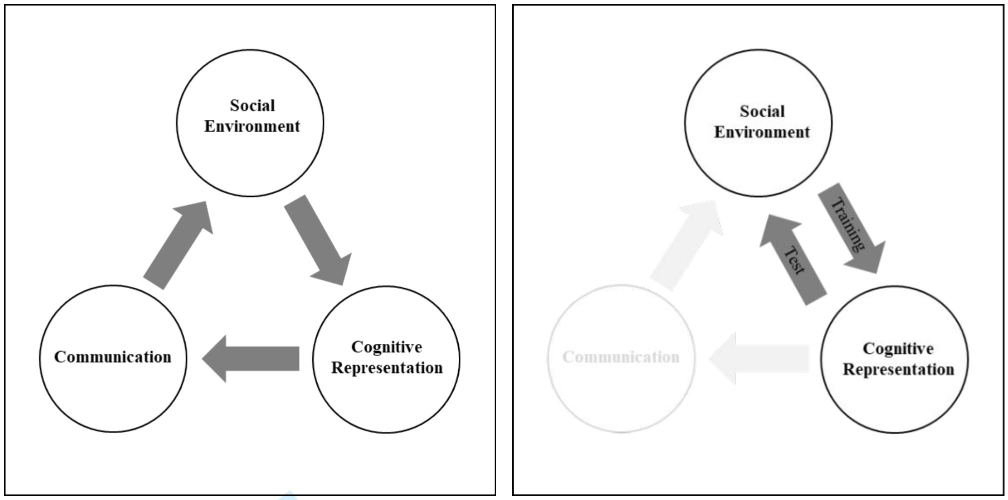


Figure 1. Schematic depictions of the proposed cycle of cultural evolution (left panel; adapted from Martin et al. 2017) and the social transmission design of the current research (right panel).

The gender stereotypes of feminine women and masculine men seem to have changed surprisingly little in the past seventy years (Donnelly & Twenge, 2017; Eagly et al., 2020; Haines et al., 2016; Lueptow & Garovich-Szabo, 2001). Evidence for the persistence of traditional gender stereotypes can be seen in explicit self-report measures of stereotype knowledge (Eagly et al., 2020; Haines et al., 2016) and in evidence from implicit reaction time measures (e.g., Nosek et al., 2009). Research using the Bem Sex Role Inventory (Bem, 1981) – a measure of people's self-identification with gender stereotypes – also reveals that women are more likely to identify with stereotypically feminine attributes and men are more likely to identify with stereotypically masculine attributes; indeed, the difference between women and men's gender stereotype self-identification has remained relatively consistent for more than forty years (Donnelly & Twenge, 2017; note that women's self-identification with masculine trait scores has increased over time, but the difference between genders has remained constant). Given that the last seventy years have seen unprecedented narrowing of gender disparities in many areas of life, such as the number of women and men in both the workplace and higher education, and apparent changes in people's attitudes towards sex and gender roles, why is it that the content of gender stereotypes remains so stubbornly impervious to change?

One reason why the content of gender stereotypes might be so pervasive is because of the biasing effect they have on cognitive representation. According to gender schema theory knowledge of gender stereotypes influences the way people cognitively assimilate social information (Bem, 1981). Thus, irrespective of whether a person endorses stereotype content, it influences the way they attend to, store in memory, and recall social information (Fyock & Stangor, 1994; Macrae & Bodenhausen, 2000; Rojahn & Pettigrew, 1992; Stangor & McMillan, 1992; Tirado, Felix-Esbri, Forn, & Sanchis-Segura, 2023). It has been shown repeatedly that people are better at remembering information that is consistent with their knowledge of stereotypes relative to information that is inconsistent with these stereotypes (for meta-analyses see Fyock & Stangor, 1994; Stangor & McMillan, 1992; although for an alternative interpretation see Rojahn & Pettigrew, 1992). Similarly, there is evidence that people experience stereotype-consistent memory intrusions that lead them to falsely believe they have encountered stereotype-consistent information that was never present (Bellezza & Bower, 1981; Lenton, Blair, & Hastie, 2001; Tirado et al., 2023). Such stereotype-consistent memory biases have the potential to influence the way individual people cognitively (mis)represent their own experience of social reality (Bem, 1981).

Any bias in memory for information consistent with gender schema (Bem, 1981) – whether it is a bias in favor of correctly recalling stereotype-consistent information (Fyock & Stangor, 1994) or a bias towards making stereotype-consistent memory intrusions (Bellezza & Bower, 1981; Lenton et al., 2001; Tirado et al., 2023) – need not be large to exert substantial influence on our collective culture and society. Even if the bias individual people exhibit is very small this can exert huge societal influence when many people possess the same or similar biases and when information is repeatedly socially transmitted (see Mesoudi, 2011). The cumulative effect of widely-shared biases – often called *cultural evolution*– is thought to be fundamental to the development of many aspects of human culture (Boyd & Richerson, 1985; Kirby, Cornish & Smith, 2008; Sperber, 1996). If enhanced memory for stereotype-consistent information is a widely shared cognitive bias, then even if this bias is subtle, even if it is not universally present in all people, and even if it is not always evident, it is plausible that the effects of this bias could accumulate to maintain societal stereotypes as information is repeatedly communicated from person to person.

Evidence from communication research suggests social information evolves to become increasingly stereotype-consistent when it is repeatedly communicated (e.g., Kashima, 2000). In dyadic conversation, people spend more time discussing stereotype-consistent information, expressing agreement with stereotype-consistent statements, and focusing questions and discussion on stereotype-consistent information (see Ruscher, 1998). This is partly because people tailor the content of their communications based on their perceptions of an audience's knowledge and expectations – a phenomenon often referred to as "audience tuning" (Echterhoff, Higgins, & Groll, 2005). These stereotype-consistent biases in communication can accumulate, with evidence that when stories are repeatedly passed from person to person, they become increasingly consistent with existing gender stereotypes (e.g., Clark & Kashima, 2007; Kashima, 2000; Lyons & Kashima, 2003). The prevailing thought is that people possess *social biases* and *communicative biases* that make them more likely to communicate stereotype-consistent information (for a review see Kashima, Klein, & Clark, 2007). However, because all communication requires recalling information from memory, if people possess *cognitive biases* that mean they are more likely to recall stereotype-consistent information (Fyock & Stangor, 1994), this might increase the likelihood of stereotype-consistent communication without the need for social and communicative biases. Thus, even in the absence of social and communicative pressure, stereotype-consistent memory bias might cause information to become increasingly stereotype-consistent as it is repeatedly passed from person to person. A central aim of the current research was therefore to establish whether, in the absence of communication bias, social information evolves to become increasingly stereotype-consistent due to memory bias in cognitive representation alone.

Recent research suggests that stereotype content might spontaneously form and change via cultural evolution without the need for social or communicative pressure (Hutchison & Martin, 2015; Hutchison et al., 2018; Martin et al., 2014; 2017). Martin et al. examined the formation and evolution of novel stereotypes using a transmission chain method, where information is repeatedly passed from person to person in a process akin to the children's game often called "Chinese Whispers" in the UK or "Telephone" in the US (Martin et al., 2014). Participants attempted to learn attributes that had been randomly associated with 'alien beings' that were individually unique but that also shared category membership (e.g., some aliens were the same shape, some were the same color). Whichever attributes one participant recalled as being associated with a particular alien were used as the basis of the training materials for the next participant in the chain of transmission. Crucially, there was no social or communicative pressure, as participants were not aware that they were part of a transmission chain. An initially random set of attributes that was difficult to remember became increasingly simplified and learnable as it passed from person to person. Even small tendencies towards structure evident in the attribute assignments made by one participant were detected and amplified by the next. Over multiple such 'generations' of transmission, a systematic stereotype-like relationship developed, with categorical features becoming so strongly associated with specific attributes that they could be used to accurately infer information about aliens that had never been seen before (e.g., by the end of one chain all green aliens were agreed to be *arrogant* and *pushy*, while red aliens were thought to be *shy*). These findings support the possibility that the formation and evolution of stereotypes is driven by shared cognitive biases, the effects of which accrue as information is repeatedly transmitted between people.

Aims and hypotheses

The overarching aim of the current research was to establish whether existing stereotypes re-emerge and persist because stereotype-consistent memory bias shapes how social information evolves as it is repeatedly passed from person to person. We know from previous research that information becomes increasingly stereotype-consistent when people are asked to communicate their knowledge to others (Clark & Kashima, 2007; Kashima, 2000; Kashima et al., 2007; Lyons & Kashima, 2003); however, it is not clear whether these effects are reliant on social and communicative biases or whether similar effects might be driven by a stereotype-consistent memory bias (Bem, 1981; Fyock & Stangor, 1994; Lenton et al., 2001). We also know from previous research that a bias towards making within-category memory errors can drive the spontaneous formation of novel stereotypes (Hutchison, 2017; Martin et al., 2014). Because micro-level gender schema are generally acknowledged to be widely shared across the population (possibly universally shared; Bem, 1981), they are an example of the kind of widely shared cognitive bias that is likely to influence macro-level cultural evolution. Is it possible, then, that a memory bias for stereotype-consistent information might lead to the re-emergence of existing stereotypes as social information is repeatedly socially transmitted?

The current research uses a "social transmission chain" design to study how stereotype driven bias in cognitive representation influences cultural evolution in a lab-based setting. In this design, people are trained on information from a novel social environment and their memory is subsequently used as the training materials for the next person in a chain. This process is repeated to create transmission chains of multiple generations, thereby mimicking how cultural knowledge is passed down through generations. Crucially, in our design participants are not asked to communicate information to other people, but rather are only asked to recall it from memory (indeed they are not even aware that they are part of a transmission chain). This allows us to investigate whether cognitive bias associated with knowledge of stereotypes alone can cause stereotypes to spontaneously re-emerge in an environment that is initially free from bias and where there is no bias from communicative pressure (see Figure 1, right panel).

We investigated whether gender stereotype-consistent memory bias shapes the cultural evolution of socially transmitted information, resulting in the spontaneous re-emergence of societal gender stereotypes that were initially absent from the social environment. Adapting the transmission chain method previously used to examine how novel stereotypes spontaneously form and evolve (Martin et al., 2014), across four experiments we examined whether people's memory for the personality attributes associated with female and male social targets became increasingly gender stereotype-consistent as information passed down transmission chains of four generations of participants. If gender schema drives a stereotype-consistent memory bias (Bem, 1981; Fyock & Stangor, 1994; Lenton et al., 2001), information should evolve to become increasingly stereotype-consistent as it is repeatedly passed from person to person.

Experiment 1

The aim of Experiment 1 was to examine whether gender stereotypes spontaneously re-emerge through a process of cultural evolution. Specifically, we wanted to examine whether repeatedly passing peoples' memories for the attributes associated with social targets down transmission chains would result in information becoming increasingly consistent with societal gender stereotypes. Participants arrived at the lab and were assigned to a novel "team" and were told that their tasks during the experiment would be to learn who was in each team[2], learn attributes associated with each person, and finally, try to remember which attributes went with which person. Participants were asked to learn personality attributes associated with sixteen social targets (8 female targets and 8 male targets). We initialized each chain by assigning to each target two stereotypically feminine attributes (e.g., *caring*), two stereotypically masculine attributes (e.g., *arrogant*) and two gender stereotype-neutral attributes (e.g., *secretive*); we refer to this initial set of target-attribute assignments as *Generation 0*. Importantly, female and male targets were assigned identical attributes at Generation 0 across the chains. During an initial training phase, the first participant in a chain (Generation 1) attempted to learn which attributes went with which target person. During a subsequent recall phase, the participant was shown each

[2] Additional conditions: The current research was part of a wider project designed to separately examine how the cumulative cultural evolution of social information is influenced by both existing stereotype knowledge *and* intergroup bias associated with novel minimal groups. To this end, participants were told that the social targets they would be learning about belonged to the same fictitious team as themselves (i.e., an in-group) or a different team to themselves (i.e., an out-group). However, as was our original intention, in the current report we only examine the intended questions and data pertaining to the re-emergence of existing gender stereotypes. We examine questions and data pertaining to the influence of minimal intergroup bias in a separate report (this will be available from an open-source repository soon).

target person again and asked to identify their attributes. The attribute-to-target pairings each person produced during the recall phase were used as the training materials for the next participant in the chain (Generation 2). The process of using the test responses from one generation as the training materials for the next was repeated three times per chain to create social transmission chains of four generations (i.e., Generation 1-4).

Because people possess gender schema that causes them to show a bias for remembering stereotype-consistent information (Bem, 1981; Fyock & Stangor, 1994; Stangor & McMillan, 1992), we hypothesised that passing initially gender-balanced social information through the memories of just four generations of participants would result in the spontaneous re-emergence of societal gender stereotypes. We expected the accumulation of stereotype-consistent information to be indexed in target-attribute pairings in two ways: first, we hypothesised that by the end of the chains the absolute frequency of stereotype-consistent target-attribute pairings would be greater than at the start of the chains (i.e., relative to the start of the chains, by Generation 4 female targets would be associated with a higher frequency of feminine attributes and male targets would be associated with a higher frequency of masculine attributes). Second, we hypothesised that by the end of the chains there would be a higher frequency of stereotype-consistent target-attribute pairings than stereotype-inconsistent target-attribute pairings (i.e., by the end of the chains feminine attributes would be paired relatively more frequently with female targets than male targets, and masculine attributes would be paired relatively more frequently with male targets than female targets).

Methods

The experiments were designed as a sub-project of a successful grant application (ES/N019121/1), in 2016, before it was commonplace for research designs to be pre-registered;

consequently, no studies reported in this manuscript were pre-registered. We report all manipulations, measures, and exclusions in these studies (although see footnote 1). All data, variable codebook, analysis code, and verbatim wording of participant instructions, have been made publicly available at the *Open Science Foundation* and can be accessed at https://osf.io/nf2pj/?view_only=4a719fd5eaae4b158385c070e2f3f181.

Participants

The participants were 64 students (52 female and 12 male), who were granted course credit in exchange for taking part in the experiment. The participants were assigned to sequential generations in one of 16 active chains and were given full instructions both verbally and onscreen. The sample size was based on previous research (Martin et al., 2014) which indicated consistent stereotype formation after four generations of transmission. *A priori* power analyses conducted using GPower (Faul, Erdfelder, Buchner, & Lang, 2009) show sufficient power (over .90 on every measure) to detect the effect sizes observed in Martin et al. (2014) with a minimum of thirteen chains of four generations each.

Materials

Social targets: Each social target was represented by one of 16 color front-facing digital headshot images of unfamiliar people (8 female and 8 male) selected from the Chicago Face Database (Ma, Correll, & Wittenbrink, 2015), cropped to a standard size of 200 x 240 pixels (1280 x 1024 screen resolution). To minimize bias associated with other social categories, all target identities were White younger adults, which reflected the racial identity and age of our participant population. We used the norms available from the Chicago Face Database to minimize between-sex differences in physical and perceived psychological characteristics (i.e., attractiveness, trustworthiness, distinctiveness, age; see Supplementary Tables A-B).

Attributes: The attributes used to describe each social target were drawn from a total pool of 48 attributes that could be used to describe a human. The chosen attributes were based on gender stereotyped and gender-neutral items that appear in the Bem Sex Role Inventory or associated synonyms (Bem, 1981)[3]; there were 16 stereotypically feminine attributes (e.g., *caring*), 16 stereotypically masculine attributes (e.g., *arrogant*) and 16 gender stereotype-neutral attributes (e.g., *secretive*). Six attributes were used to describe each target.

Procedure

Generating the initial system (Generation 0): We created Generation 0 by assigning six attributes to each target in a pseudo-random way that minimized categorical structure associated with target sex. Each social target was assigned two stereotypically feminine attributes (e.g., *caring*), two stereotypically masculine attributes (e.g., *arrogant*), and two relatively gender stereotype-neutral attributes (e.g., *secretive*). Importantly, female and male targets were each collectively assigned identical attributes at Generation 0 across the chains. The specific assignment of attributes to targets was counterbalanced across chains.

Stage 1. Minimal group induction: Participants were assigned to one of two teams within the experiment and were given two minutes to study two office layouts, which included images of each team member and where they supposedly sat. Participants were then tested on their knowledge of the members of both teams by being asked to remember where each person sat and via a simple face recognition test.

[3] We collected femininity-masculinity ratings for each attribute (Likert scale 1 = most feminine – 5 = most masculine) using separate pilot samples of equivalent undergraduates (see Supplementary Table C). Subsequent analysis of these ratings revealed the mean ratings of feminine, masculine, and neutral items were all significantly different from one another (Supplementary Table D).

Stage 2. Attribute training: Participants were told that they would now be presented with personality attributes that described each of their colleagues, and that their task was to form an impression of each person and to remember their attributes. Each trial lasted 20-seconds and comprised the presentation of a face image and six personality attributes that described the person. To aid learning, participants viewed each target and their associated information three times (i.e., 48 trials in total). The order of target presentation was randomized across trials.

Stage 3. Attribute memory test: Participants were asked to try to remember which attributes went with which targets in the attribute training stage (stage 2). Each trial comprised the presentation of a single target face below a list of the entire pool of 48 possible attributes. Participants completed 16 trials, one per target person, with the trial order randomized. Each target remained onscreen until six attributes had been chosen, with participants encouraged to make their best guess if they were unsure about which attributes to choose.

Social transmission of information (Generations 1-4): Following each participant's completion of the experiment, the attributes they assigned during the memory test stage (Stage 3) were transmitted as the materials for the attribute training stage of the next participant (Stage 2). The process of transmitting the memory test responses from one participant as the impression formation materials for the next was repeated to create continuous transmission chains of four generations.

Dependent measures and analysis strategy

Frequency of stereotype-consistent and stereotype-inconsistent target-attribute pairings over time: To establish whether information became increasingly stereotype-consistent as it passed down the chains we analysed the mean proportion target-attribute frequencies with 2 (Generation: G1 vs. G4) X 2 (Target Sex: female targets vs. male targets) X 2 (Attribute Type:

stereotype-consistent vs. stereotype-inconsistent)[4] repeated measures Analysis of Variance (ANOVA). If information became increasingly stereotype-consistent over time we expected to find a significant Generation X Attribute Type interaction driven by a significant increase in stereotype-consistent target-attribute pairings but not stereotype-inconsistent target attribute pairings.

Experiment 1 Results

Frequency of stereotype-consistent and stereotype-inconsistent target-attribute pairings over time: The only significant effect to emerge from the analysis was a main effect of Generation [$F(1, 15) = 22.41$, $p < .001$; $\eta^2 = .599$; see Supplementary Table 1 for all cell values)]. However, this effect was included in a non-significant trend towards the predicted interaction of Generation X Attribute Type [$F(1, 15) = 3.99$, $p = .064$; $\eta^2 = .210$], which is explored in more detail below. There was no evidence of any other significant main effects or interactions (all Fs < .174, all ps > .120).

Given its central importance to the current research, we further explored the non-significant trend towards the predicted Generation X Attribute Type interaction; we did this by examining the pairwise comparisons for stereotype-consistent and stereotype-inconsistent target-attribute pairings both within and between each Generation (Figure 2[5]). As predicted, stereotype-consistent target-attribute pairings increased significantly between G1 (*M prop* = .31; 95% CI

[4] Because the data for the dependent measure are proportions of the whole, it is not appropriate to run an omnibus ANOVA with a factor of Attribute Type that has three levels (i.e., 3 (Attribute Type: stereotype-consistent vs. stereotype-inconsistent vs. stereotype neutral); doing so merely results in many comparisons between levels that have identical means and zero variance.

[5] While our statistical analyses focus on the key difference in responses between the start and end of the chain (G1 vs. G4), we have graphically depicted the data for all four generations in the figures to illustrate the incremental nature of the cultural evolution process.

[.27, .35]) and G4 (*M prop* = .39; 95% CI [.34, .43]; *M diff* = .08 (95% CI [.03, .13]; *p* = .003; *d* = .88, 95% CI [.29, 1.45]); there was no change in the proportion of stereotype-inconsistent target-attribute pairings between G1 (*M prop* = .36; 95% CI [.31, .41]) and G4 (*M prop* = .36; 95% CI [.31, .40]; *M diff* = .003 (95% CI [-.04, .04]; *p* = .904; *d* = .03, 95% CI [-.46, .52]). There was no difference in the frequency of stereotype-consistent and stereotype-inconsistent attributes at G1 (*M diff* = .05; 95% CI [-.03, .14]; *p* = .209; *d* = .33, 95% CI [-.18, .83]); contrary to our predictions, there was also no difference in the frequency of stereotype-consistent and stereotype-inconsistent attributes at G4 (*M diff* = .03; 95% CI [-.05, .11]; *p* = .463; *d* = .19, 95% CI [-.31, .68]).

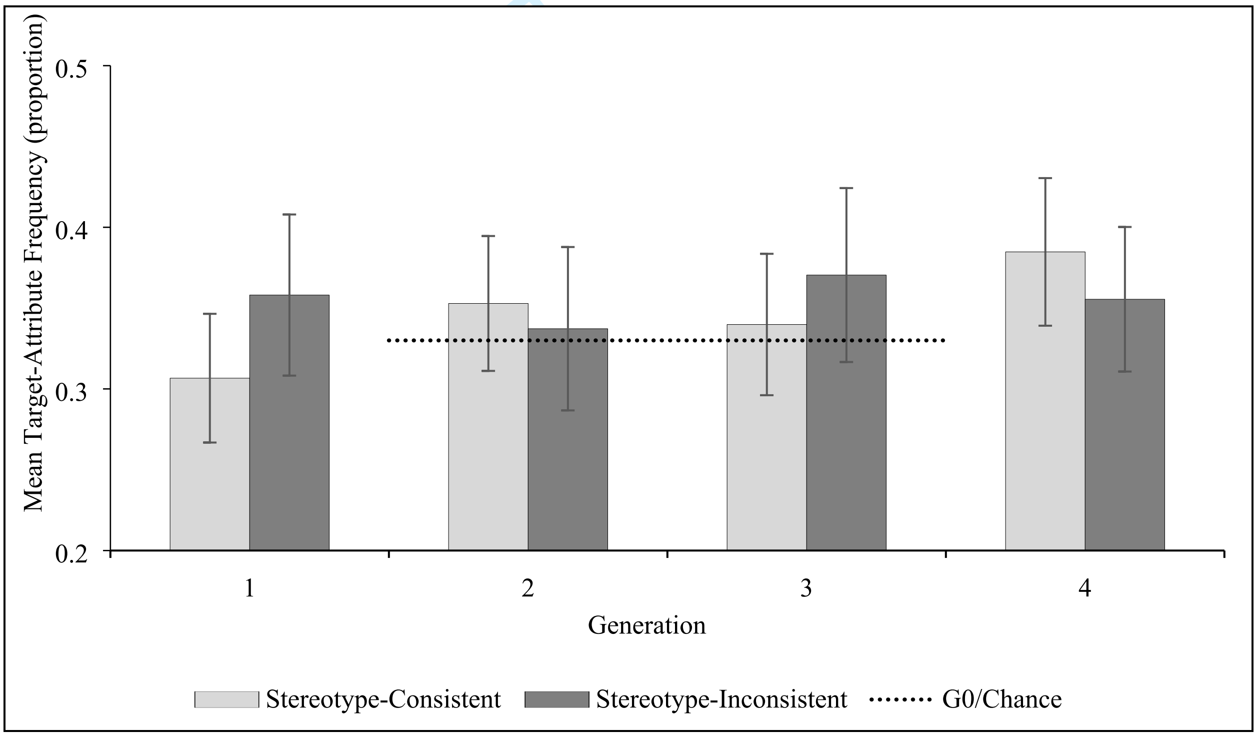


Figure 2. Expt. 1 frequency of target-attribute pairings by Generation and Attribute Type. Error bars represent 95% confidence intervals.

Discussion – Experiment 1

The aim of Experiment 1 was to examine whether gender stereotypes spontaneously re-emerge through a process of cultural evolution – we found only limited evidence that this was the case. As predicted, there was evidence that stereotype-consistent information accumulated as it passed down the chains, with significantly more stereotype-consistent target attributes at the end of the chains than at the start. However, contrary to our prediction, there was no evidence that by the end of the chains there would be more stereotype-consistent target-attribute pairings than stereotype-inconsistent target-attribute pairings.

While we found some evidence that information evolved to become increasingly consistent with gender stereotypes as it passed down the chains, there are several reasons why this might not have culminated in the spontaneous re-emergence of gender stereotypes by the end of the chains. Directional cultural evolution of information depends on the strength of biases of individual learners, how many people share these biases, and how often information is communicated (Hutchison et al., 2018; Martin et al., 2017; Mesoudi, 2011). It is likely that the current experimental context resulted in an individual memory bias towards stereotype-consistent information that was smaller than we anticipated (Fyock & Stangor, 1994; Lenton et al., 2001). Given the frequency of stereotype-consistent target-attribute pairings was on an increasing trend, with substantial increases in stereotype-consistent pairings between Generation 3 and Generation 4, it is likely this level of bias might result in stereotype re-emergence with more generations of participants per chain. However, rather than examining the cumulative effects of weak bias over a greater number of generations, an alternative approach is to examine whether stronger bias might result in the re-emergence of existing stereotypes over the same number of generations – we explored whether a stronger gender stereotype-consistent memory bias would lead to the re-emergence of gender stereotypes in Experiment 2.

Experiment 2

The aim of Experiment 2 was to induce a stronger stereotype-consistent memory bias, by introducing a novel gendered occupational context, to examine whether gender stereotypes spontaneously re-emerge through a process of cultural evolution. Gender-based occupational stereotypes are very prevalent, with jobs requiring communal skills more likely to be perceived as feminine and undertaken by women (e.g., nurses, secretaries, elementary school teachers), and jobs requiring agentic skills more likely to be perceived as masculine and undertaken by men (e.g., engineers, computer programmers, surgeons; Atli, 2017; Matheus & Quinn, 2017; Treleaven, 2015; White & White, 2006). Recent research suggests gender schema of occupations can lead to stereotype-consistent memory bias, with people more likely to correctly and falsely recall information that is consistent with occupational gender stereotypes (Tirado et al., 2023). It is therefore possible the presence of gendered occupational context might lead to a stronger gender stereotype-consistent memory bias, which in turn leads to the re-emergence of gender stereotypes over time. Thus, in Experiment 2, we examined whether a feminine-stereotyped social environment influenced the maintenance of gender stereotypes, by describing targets as belonging to one of two occupational teams with novel feminine stereotyped names (i.e., *community outreach team* or *social resources team*). We chose to use novel occupational titles, which were aligned with gender stereotypes but which were not themselves associated with strong occupational stereotypes. This is because we were concerned that using strongly gendered real-world roles (e.g., nurse, surgeon) would result in the re-emergence of stereotypes associated with those occupations/settings rather than gender per se. Our predictions were the same as those in Expt. 1.

Methods

Participants. The participants were 64 students (60 female and 4 male) from the University of Aberdeen, who were granted course credit in exchange for taking part in the experiment.

Materials. The materials used were identical to those of Experiment 1, except that references to "your team" and "the other team" were replaced with "community outreach team" and "social resources team" respectively; pilot testing indicated that these fictitious job titles were strongly associated with feminine stereotyped attributes (see Supplementary Tables E-F).

Procedure. The procedure was the same as that of Experiment 1.

Dependent Measures and Analyses. The dependent measures and analyses were identical to those used in Experiment 1.

Experiment 2 Results

Frequency of stereotype-consistent and stereotype-inconsistent target-attribute pairings over time: We analysed the data for the mean proportional frequency with which female and male targets were paired with feminine and masculine stereotyped attributes at the start and end of the chains using a 2 (Generation: G1 vs. G4) X 2 (Target Sex: female targets vs. male targets) X 2 (Attribute Type: stereotype-consistent vs. stereotype-inconsistent) repeated measures ANOVA (see Supplementary Table 2 for all cell values). The analysis revealed a main effect of Attribute Type [$F(1, 15) = 7.35$, $p = .016$; $\eta^2 = .329$], but this was subsumed by the predicted interaction of Generation X Attribute Type [$F(1, 15) = 5.20$, $p = .038$; $\eta^2 = .257$], which is explored in more detail below. There was no evidence of any other significant main effects or interactions (all Fs $< .467$, all ps $> .504$).

We explored the predicted Generation X Attribute Type interaction by examining the pairwise comparisons for stereotype-consistent and stereotype-inconsistent target-attribute pairings both within and between Generation 1 and Generation 4 (Figure 3). As predicted, stereotype-consistent target-attribute pairings increased significantly between G1 (*M prop* = .36; 95% CI [.33, .39]) and G4 (*M prop* = .41; 95% CI [.36, .45]; *M diff* = .05 (95% CI [.01, .09]; *p* = .018; *d* = .66, 95% CI [.11, 1.19]); there was no change in the proportion of stereotype-inconsistent target-attribute pairings between G1 (*M prop* = .34; 95% CI [.30, .38]) and G4 (*M prop* = .29; 95% CI [.26, .34]; *M diff* = .04 (95% CI [-.01, .10]; *p* = .109; *d* = .43, 95% CI [-.09, .93]). There was no difference in the proportion of stereotype-consistent attributes and stereotype-inconsistent attributes used at G1 (*M diff* = .02; 95% CI [-.05, .08]; *p* = .608; *d* = .13, 95% CI [-.36, .62]); however, as predicted there was a significantly higher proportion of stereotype-consistent than stereotype-inconsistent target-attribute pairings at G4 (*M diff* = .11; 95% CI [.04, .17]; *p* = .003; *d* = .87, 95% CI [.28, 1.44]).

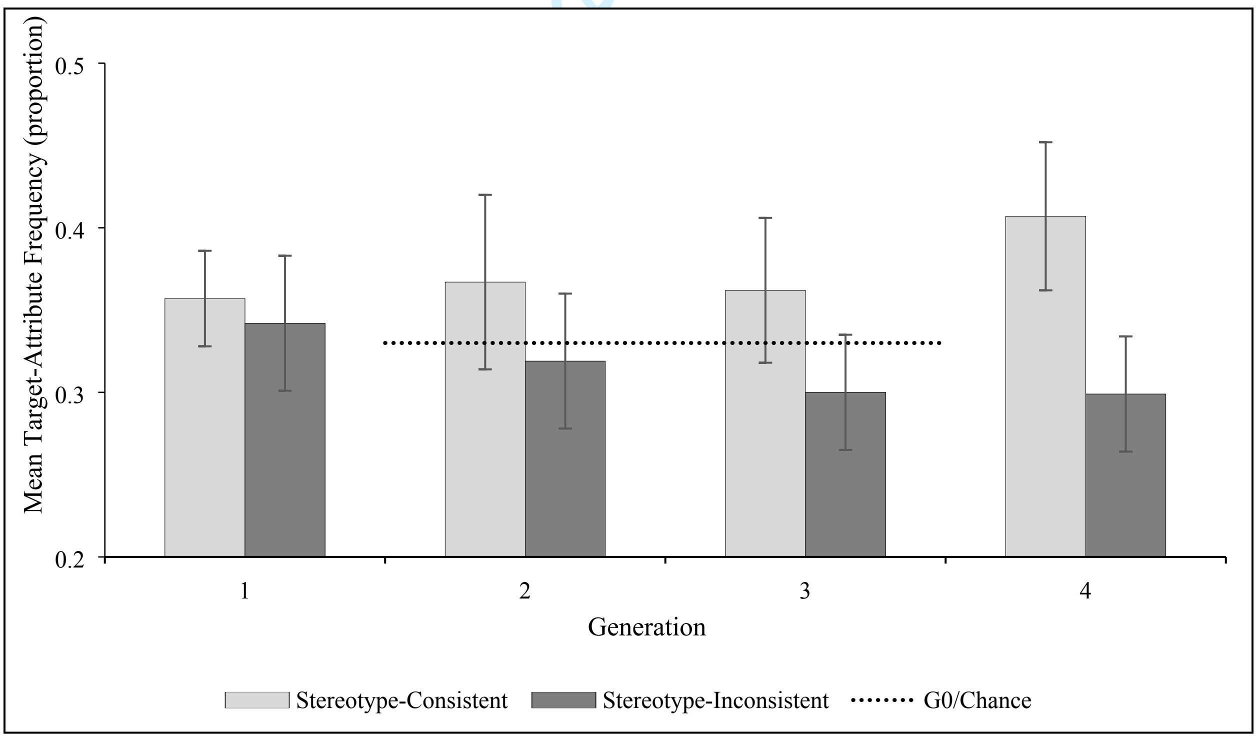


Figure 3. Expt. 2 frequency of target-attribute pairings by Generation and Attribute Type. Error bars represent 95% confidence intervals.

Discussion – Experiment 2

The aim of Experiment 2 was to induce a stronger stereotype-consistent memory bias, by introducing a gendered occupational context, to examine whether gender stereotypes spontaneously re-emerge through a process of cultural evolution – we found evidence that this was the case. As predicted, and as in Expt.1, there was evidence that stereotype-consistent information accumulated as it passed down the chains, with significantly more stereotype-consistent target attributes at the end of the chains than at the start. However, unlike in Expt. 1, there was also evidence to support our prediction that by the end of the chains there would be more stereotype-consistent target-attribute pairings than stereotype-inconsistent target-attribute pairings. Given that in a stereotypically feminine occupational context female targets became increasingly associated with feminine attributes and male targets became increasingly associated with masculine attributes, one would expect the same pattern in a masculine context, which provides an opportunity to test the replicability of this finding. We explored this possibility in Experiment 3.

Experiment 3

The aim of Experiment 3 was to examine whether gender stereotypes spontaneously re-emerge through a process of cultural evolution in the context of a masculine occupational context. As was the case in Expt. 2, we created a gendered social context by describing targets as belonging to one of two novel teams, in this case with stereotypically masculine sounding names (i.e., *mobility analyst team* or *product factors team*). Our predictions were the same as those for Expt. 1 and Expt. 2.

Methods

Participants. The participants were 64 undergraduate students (32 male and 32 female)[6] from the University of Aberdeen, who were granted course credit in exchange for taking part in the experiment.

Materials. The materials used were identical to those of Experiment 2, except references to the relatively feminine team names were replaced with the masculine team names of "mobility analysts" and "product factors"; pilot testing indicated that these fictious job titles were strongly associated with masculine stereotyped attributes (see Supplementary Tables E-F).

Procedure. The procedure was the same as that of Experiment 2.

Dependent Measures and Analyses. The dependent measures and analyses were identical to those used in Experiment 1 and 2.

Experiment 3 Results

Frequency of stereotype-consistent and stereotype-inconsistent target-attribute pairings over time: We analysed the data for the mean proportional frequency with which female and male targets were paired with feminine and masculine stereotyped attributes at the start and end of the chains using a 2 (Generation: G1 vs. G4) X 2 (Target Sex: female targets vs. male targets) X 2 (Attribute Type: stereotype-consistent vs. stereotype-inconsistent) repeated measures ANOVA (see Supplementary Table 3 for all cell values). The analysis revealed a main effect of Attribute Type [$F(1, 15) = 12.30$, $p = .003$; $\eta^2 = .451$], but this was subsumed by the predicted interaction of Generation X Attribute Type [$F(1, 15) = 10.04$, $p = .006$; $\eta^2 = .401$], which is explored in more detail below. There were also several unpredicted significant effects and trends, including a significant interaction of Target Sex and

[6] In Expt. 3, we recruited an equal number of female and male participants to minimise the possibility that the effects were being driven by participant sex or gender. Participants were tested in within participant gender chains (i.e., chains of four generations of women participants or chains of four generations of men participants). Preliminary exploratory analyses revealed there to be no significant effect of chain gender (although these analyses were underpowered).

Attribute Type [$F(1, 15) = 7.96$, $p = .013$; $\eta^2 = .347$], which is explored in more detail below, a non-significant trend towards a main effect of Target Sex [$F(1, 15) = 4.24$, $p = .057$; $\eta^2 = .220$], and a non-significant trend towards a three-way interaction [$F(1, 15) = 3.24$, $p = .092$; $\eta^2 = .178$]. There was no evidence of any other significant main effects or interactions (all *F*s < 2.25, all *p*s > .154).

We explored the predicted Generation X Attribute Type interaction by examining the pairwise comparisons for stereotype-consistent and stereotype-inconsistent target-attribute pairings both within and between Generation 1 and Generation 4 (Figure 4). As predicted, stereotype-consistent target-attribute pairings increased significantly between G1 (*M prop* = .36; 95% CI [.31, .40]) and G4 (*M prop* = .44; 95% CI [.39, .49]; *M diff* = .08; 95% CI [.03, .13]; *p* = .005; *d* = .81, 95% CI [.23, 1.37]); conversely, stereotype-inconsistent target-attribute pairings decreased significantly between G1 (*M prop* = .32; 95% CI [.29, .35]) and G4 (*M prop* = .27; 95% CI [.22, .31]; *M diff* = .05 (95% CI [.01, .10]; *p* = .026; *d* = .62, 95% CI [.07, 1.15]). There was no difference in the frequency of stereotype-consistent and stereotype-inconsistent attributes used at G1 (*M diff* = .04; 95% CI [-.03, .11]; *p* = .287; *d* = .28, 95% CI [.23, .77]); however, as predicted, there was a significantly higher frequency of stereotype-consistent than stereotype-inconsistent target-attribute pairings at G4 (*M diff* = .17; 95% CI [.09, .25]; *p* < .001; *d* = 1.09, 95% CI [.45, 1.70]).

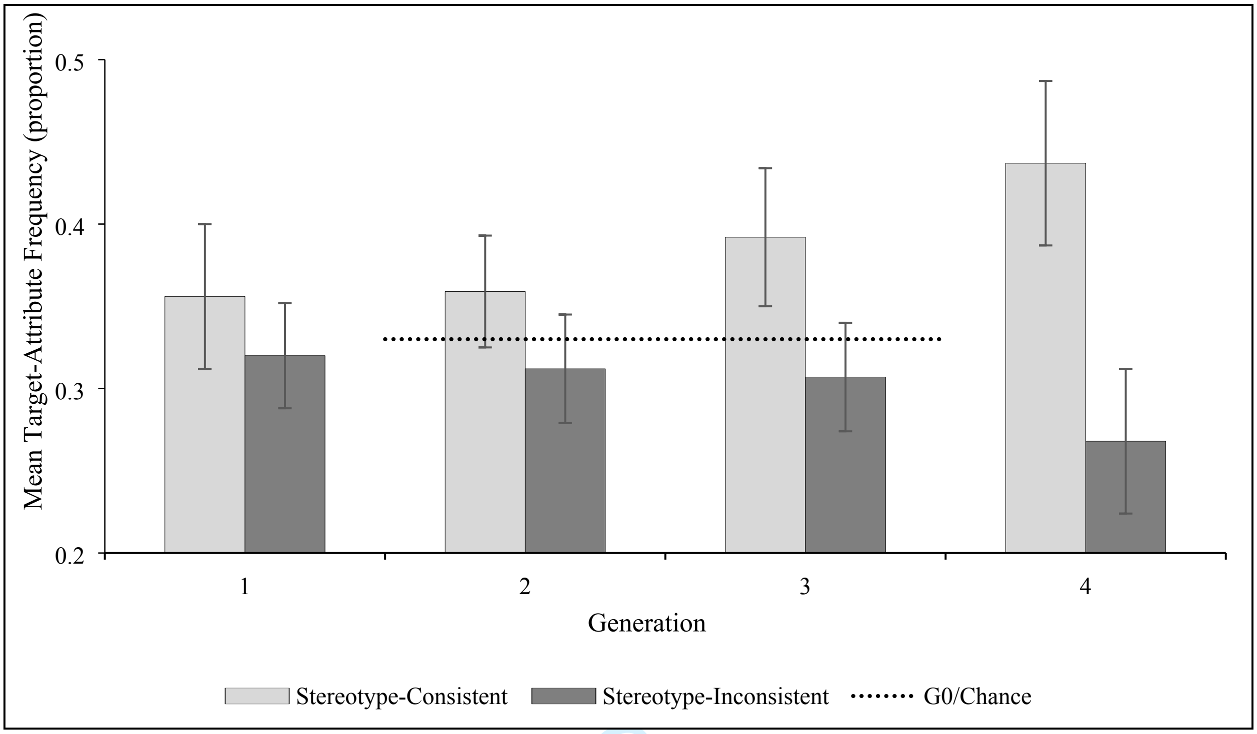


Figure 4. Expt. 3 frequency of target-attribute pairings by Generation and Attribute Type. Error bars represent 95% confidence intervals.

Given its potential theoretical interest, we also explored the unexpected Target Sex X Attribute Type interaction, by inspecting the pairwise comparisons for stereotype-consistent and stereotype-inconsistent target-attribute pairings between female and male targets. Stereotype-consistent pairings were significantly more frequent for male targets (*M prop* = .44; 95% CI [.39, .49]) than for female targets (*M prop* = .36; 95% CI [.32, .40]; *M diff* = .08; 95% CI [.03, .14]; $p = .004$; $d = .86$, 95% CI [.27, 1.42]); whereas, there was a non-significant trend in the opposite direction for stereotype-inconsistent pairings, which were numerically more frequent for female targets (*M prop* = .32; 95% CI [.28, .36]) than for male targets (*M prop* = .27; 95% CI [.23, .31]; *M diff* = .05; 95% CI [-.01, .01]; $p = .087$; $d = .46$, 95% CI [-.07, .97]).

Discussion – Experiment 3

The findings from Expt. 3 further support and extend those from Expt. 1 and Expt. 2. There was further evidence that memory bias for stereotype-consistent information accumulated as it passed down the chains. As in both Expt. 1 and Expt. 2, there was evidence that stereotype-consistent information accumulated as it passed down the chains, with significantly more stereotype-consistent target attributes at the end of the chains than at the start. As in Expt. 2, there was also evidence to support our prediction that by the end of the chains there would be more stereotype-consistent target-attribute pairings than stereotype-inconsistent target-attribute pairings. Unlike in Expt. 1 and Expt. 2, there was also evidence that size of the stereotype maintenance differed dependent on target gender, with male targets becoming more strongly associated with stereotype-consistent attributes than female targets.

Cross-Experiment Analysis (Expt. 1-3)

We ran a series of cross-experiment analyses to establish whether there was statistical support for the apparent differences between the experimental contexts from Expt. 1-3 and to further illuminate the mechanism underlying the effects. The results from Expt.1-3 support the idea that gender stereotypes can re-emerge, through a process of cultural evolution, as social information is repeatedly passed from person to person. However, the pattern of results across the experiments also seems to indicate that the extent of stereotype re-emergence differs dependent on target gender and social context. Based on these emergent patterns in the data, we ran exploratory cross-experiment analyses to test whether there was statistical support for these trends. First, we wanted to determine whether there was greater evidence of stereotype maintenance in more gendered social contexts (Expt. 2 & 3) than in a less gendered social context (Expt. 1). Second, we wanted to establish whether there was greater evidence of stereotype maintenance for male targets than for female targets.

We also used the cross-experiment analyses as an opportunity to explore two additional questions that arose during the peer review process, specifically 1) whether the observed effects were driven by better memory for stereotype-consistent information or a bias towards making stereotype-consistent memory intrusions, and 2) whether the size of stereotype-consistent memory bias was the same across the generations or whether it changes as the social environment becomes increasingly stereotype-consistent.

Results – Cross-Experiment Analysis (Expt. 1-3)

We addressed the first three aims of the cross-experiment analysis by analysing the mean proportional frequency with which female and male targets were correctly and incorrectly paired with feminine and masculine stereotyped attributes at the start and end of the chains using a 3 (Experiment Context: neutral vs. feminine vs. masculine) X 2 (Generation: G1 vs. G4) X 2 (Target Sex: female targets vs. male targets) X 2 (Attribute Type: stereotype-consistent vs. stereotype-inconsistent) X 2 (Accuracy: correct vs. intrusions) mixed ANOVA with Experiment Context as the only between-subjects factor (see Supplementary Table 4 for all cell values). We only describe the effects of central theoretical interest, but a full description of all effects can be found in Supplementary Table 5.

Is the extent of stereotype maintenance influenced by social context? Supporting the idea that stereotype maintenance differed in different social contexts, there was evidence of a significant Experiment X Attribute Type interaction [$F(2, 45) = 4.01$, $p = .025$; $\eta^2 = .151$; see Figure 5]. Inspection of the pairwise comparisons revealed that the frequency of stereotype-consistent target-attribute pairings was significantly higher in the masculine context (*M prop* = .40; 95% CI [.36, .43]) than in the neutral context (*M prop* = .35, 95% CI [.31, .38]; *M diff* = .05, 95% CI [.003, .10], $p = .038$; $d = .72$, 95% CI [.001, 1.43], but that there was no difference between the masculine context and the feminine context (*M prop* = .38, 95% CI [.35, .42]; *M diff* = .01, 95% CI [-.03, .06], $p = .549$; $d = .21$, 95% CI [-.48, .91]), or between the feminine context and the neutral context (*M diff* = .04, 95% CI [-.01, .08], $p = .132$; $d = .57$, 95% CI [-.14, 1.28]). The frequency of stereotype-inconsistent target-attribute pairings was significantly lower in the masculine context (*M prop* = .29; 95% CI [.26, .33]) than in the neutral context (*M prop* = .36, 95% CI [.33, .39]; *M diff* = .06, 95% CI [.02, .11], $p = .007$; $d = .92$, 95% CI [.19, 1.67], but there was no difference between the masculine context and the feminine context (*M prop* = .32, 95% CI [.29, .35]; *M diff* = .03, 95% CI [-.02, .07], $p = .236$; $d = .49$, 95% CI [-.22, 1.19]), or between the feminine context and the neutral context (*M diff* = .04, 95% CI [-.01, .08], $p = .115$; $d = .54$, 95% CI [-.17, 1.25]).

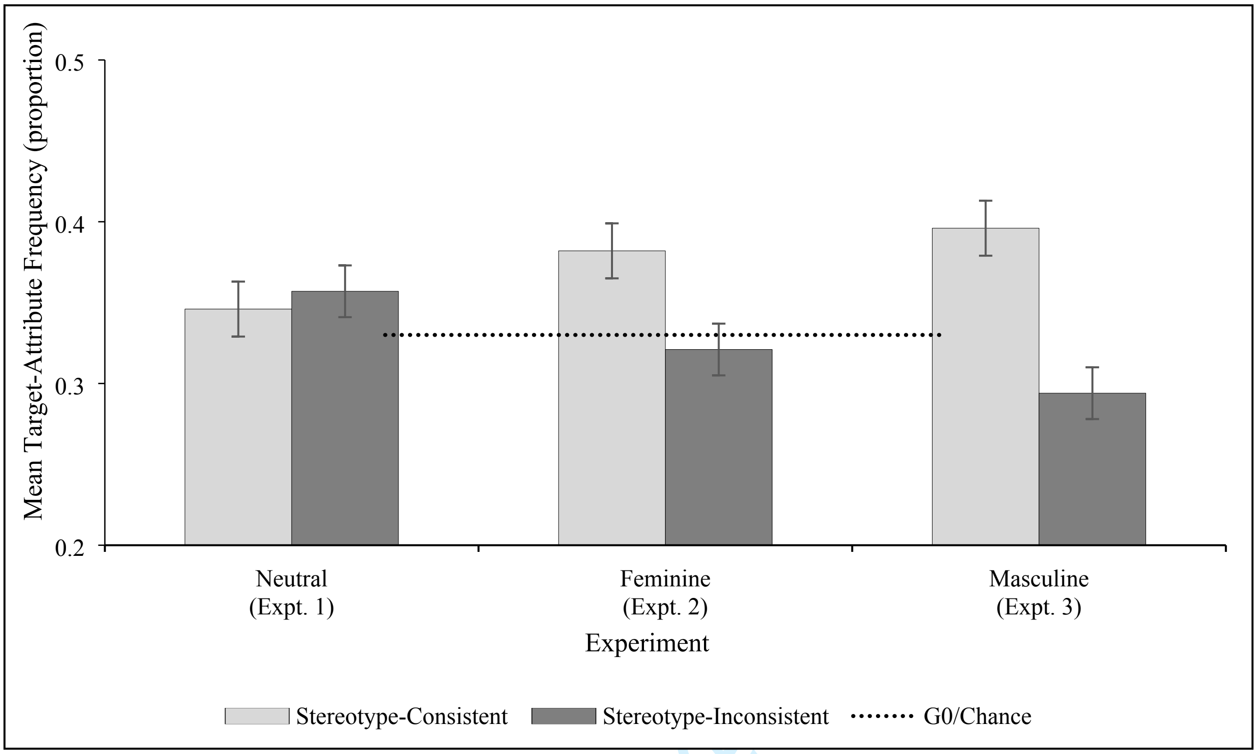


Figure 5. Expt. 1-3 frequency of target-attribute pairings by Experiment and Attribute Type. Error bars represent 95% confidence intervals.

Is there greater evidence of stereotype maintenance for male targets than female targets? Supporting the possibility that there was greater stereotype maintenance for male targets there was evidence of a Target Sex X Attribute Type interaction [$F(1, 45) = 6.29$, $p = .016$; $\eta^2 = .123$]. Stereotype-consistent attributes were paired more frequently with male targets (*M prop* = .40; 95% CI [.37, .43]) than female targets (*M prop* = .35; 95% CI [.33, .37]; *M diff* = .05; 95% CI [.01, .09], $p = .010$; $d = .39$, 95% CI [.09, .68]). Conversely, stereotype-inconsistent attributes were paired more frequently with female targets (*M prop* = .34; 95% CI [.32, .37]) than male targets (*M prop* = .31; 95% CI [.28, .33]; *M diff* = .04; 95% CI [.001, .07]; $p = .044$; $d = .30$, 95% CI [.01, .59]).

Are the effects driven by better memory for stereotype-consistent information or a bias towards making stereotype-consistent memory errors? There was evidence of a significant Generation X Accuracy X Attribute Type interaction [$F(1, 45) = 5.84$, $p = .020$; $\eta^2 = .115$; see Figure 6]. Examination of the proportion of correct responses indicates that a memory advantage for stereotype-consistent correct responses emerged over time. At G1 there was no difference in the proportion of correct responses that were stereotype-consistent (*M prop* = .08; 95% CI [.07, .09]) relative to stereotype-inconsistent (*M prop* = .08; 95% CI [.07, .08]; *M diff* = .006; 95% CI [-.004, .017], $p < .223$; $d = .18$, 95% CI [-.11, .46]); however, by G4 there were significantly more correct responses that were stereotype-consistent (*M prop* = .16; 95% CI [.14, .17]) than stereotype-inconsistent (*M prop* = .13; 95% CI [.11, .15]; *M diff* = .03; 95% CI [.003, .05], $p < .032$; $d = .31$, 95% CI [.02, .60]). Although between G1 and G4 there was a significant increase in both stereotype-consistent correct responses (*M diff* = .07; 95% CI [.06, .09], $p < .001$; $d = 1.21$, 95% CI [.84, 1.59]) and stereotype-inconsistent correct responses (*M diff* = .05; 95% CI [.03, .07], $p < .001$; $d = .76$, 95% CI [.44, 1.08]), this increase was larger for stereotype-consistent correct responses.

Examination of the proportion of intrusions indicates that a bias towards stereotype-consistent intrusions emerged over time. At G1 there was no difference in the proportion of intrusions that were stereotype-consistent (*M prop* = .26; 95% CI [.24, .28]) relative to stereotype-inconsistent (*M prop* = .27; 95% CI [.24, .29]; *M diff* = .006; 95% CI [-.03, .04], $p < .743$; $d = .05$, 95% CI [-.24, 1.33]); however, by G4 there were significantly more intrusions that were stereotype-consistent (*M prop* = .26; 95% CI [.23, .28]) than stereotype-inconsistent (*M prop* = .18; 95% CI [.17, .20]; *M diff* = .08; 95% CI [.05, .10], $p < .001$; $d = .75$, 95% CI [.43, 1.07]). The emergent bias towards stereotype-consistent intrusions occurred because the proportion of stereotype-consistent intrusions did not change between G1 and G4 (*M diff* = .004; 95% CI [-.02, .03], $p = .749$; $d = .04$, 95% CI [-.24, .33]), whereas there was a significant decrease in stereotype-inconsistent intrusions between G1 and G4 (*M diff* = .08; 95% CI [-.02, .03], $p < .001$; $d = 1.00$, 95% CI [.65, 1.34]).

When the proportion of correct responses and intrusions are combined the effects further support those found in Expt. 1-3. At G1 there was no difference in the frequency of target-attribute pairings that were stereotype-consistent (*M prop* = .34; 95% CI [.32, .36]) relative to stereotype-inconsistent (*M prop* = .34; 95% CI [.32, .36]; *M diff* = .00; 95% CI [-.04, .40], $p = .991$; $d = .002$, 95% CI [-.28, .28]). As predicted, at G4 there was a significantly higher frequency of target-attribute pairings that were stereotype-consistent (*M prop* = .41; 95% CI [.38, .44]) than stereotype-inconsistent (*M prop* = .31; 95% CI [.29, .33]; *M diff* = .10; 95% CI [.06, .14]; $p < .001$; $d = .66$, 95% CI [.35, .97]). The pairwise comparisons also revealed that there were significantly more intrusions than correct responses for each combination of Attribute Type and Generation (all *p*s < .001).

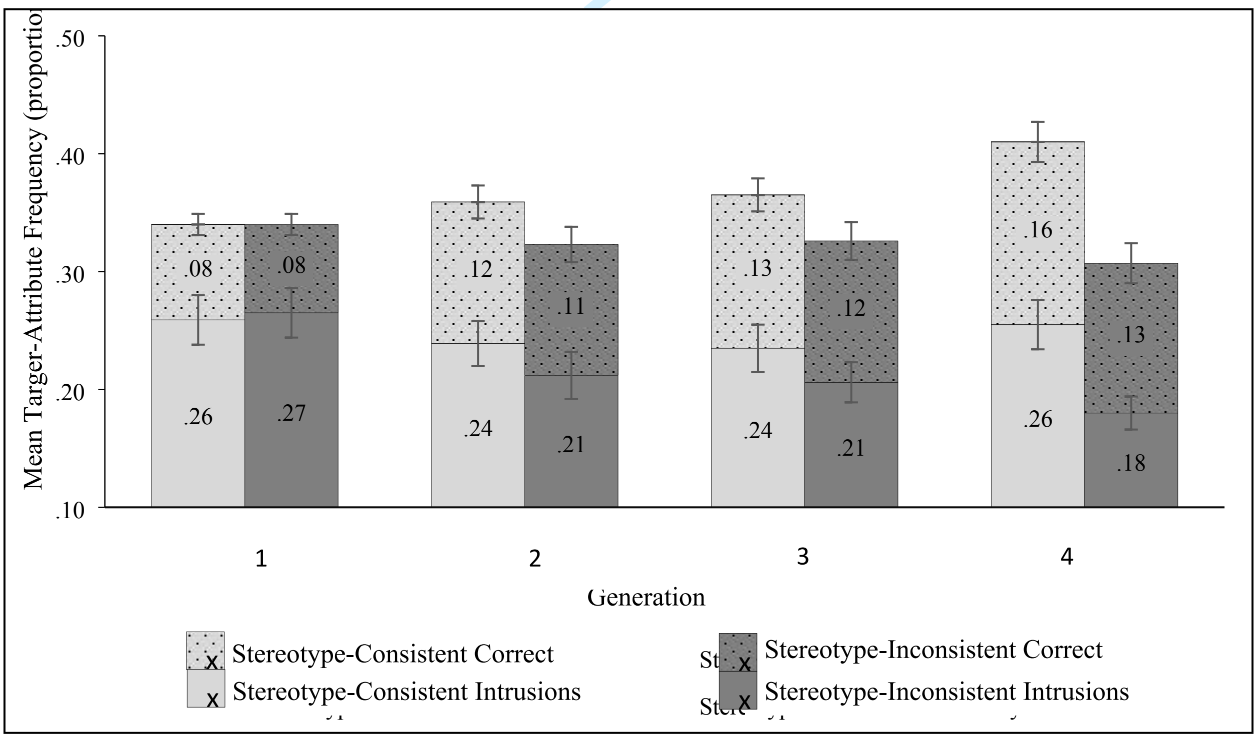


Figure 6. Expt. 1-3 frequency of target-attribute pairings by Generation, Attribute Type, and Accuracy. Numbers in each bar represent the proportion of responses for each level and error bars denote 95% confidence intervals for each level.

Is the size of stereotype-consistent memory bias the same across the generations or does it change as the social environment becomes increasingly stereotype-consistent? We addressed the final aim of the cross-experiment analysis by investigating the level of stereotype-consistent bias added by each generation; we did this subtracting the proportion of stereotype-consistent attribute pairings from the training phase from the proportion of stereotype-consistent attributes pairings from the test phase for each Generation. We then analysed the stereotype-consistent memory bias within each Generation using a 3(Experiment Context: neutral vs. feminine vs. masculine) X 2(Generation: G1 vs. G4) X 2(Target Sex: female targets vs. male targets) X 2(Attribute Type: stereotype-consistent vs. stereotype-inconsistent) mixed ANOVA with Experiment Context as the only between-subjects factor (see Figure 6). The analysis revealed a main effect of Target Sex [$F(1, 45) = 4.21$, $p = .046$; $\eta^2 = .086$], with a higher stereotype-consistent bias for male targets (*M prop* = .014; 95% CI [.006, .022]) than female targets (*M prop* = .014; 95% CI [-.001, .013]). There was also a significant main effect of Attribute Type [$F(1, 45) = 6.15$, $p = .017$; $\eta^2 = .120$], but this was subsumed by a significant interaction of Generation X Attribute Type [$F(1, 45) = 4.14$, $p = .048$; $\eta^2 = .084$], which is explored below. There was no evidence of any other significant main effects or interactions (all *F*s < 2.31, all *p*s > .111).

The Generation X Attribute Type interaction can be seen in Figure 7. Inspection of the pairwise comparisons revealed a significant increase in the frequency of stereotype-consistent bias between G1 (*M prop* = .007; 95% CI [-.014, .028]) and G4 (*M prop* = .045, 95% CI [.021, .069]; *M diff* = .038, 95% CI [.006, .070], $p = .021$; $d = .35$, 95% CI [.052, .634], but no change in the frequency of stereotype-inconsistent bias between G1 (*M prop* = .007; 95% CI [-.016, .029]) and G4 (*M prop* = -.18, 95% CI [-.003, .040]; *M diff* = .025, 95% CI [-.011, .060], $p = .167$; $d = .21$, 95% CI [-.081, .491]. There was no difference between stereotype-consistent bias and stereotype-inconsistent bias at G1 (*M diff* < .001, 95% CI [-.40, .040], $p = .991$; $d = .002$, 95% CI [-.28, .28]), whereas there was a significant difference between stereotype-consistent bias and stereotype-inconsistent bias at G4 (*M diff* = .063, 95% CI [.022, .104], $p = .003$; $d = .45$, 95% CI [.15, .75]).

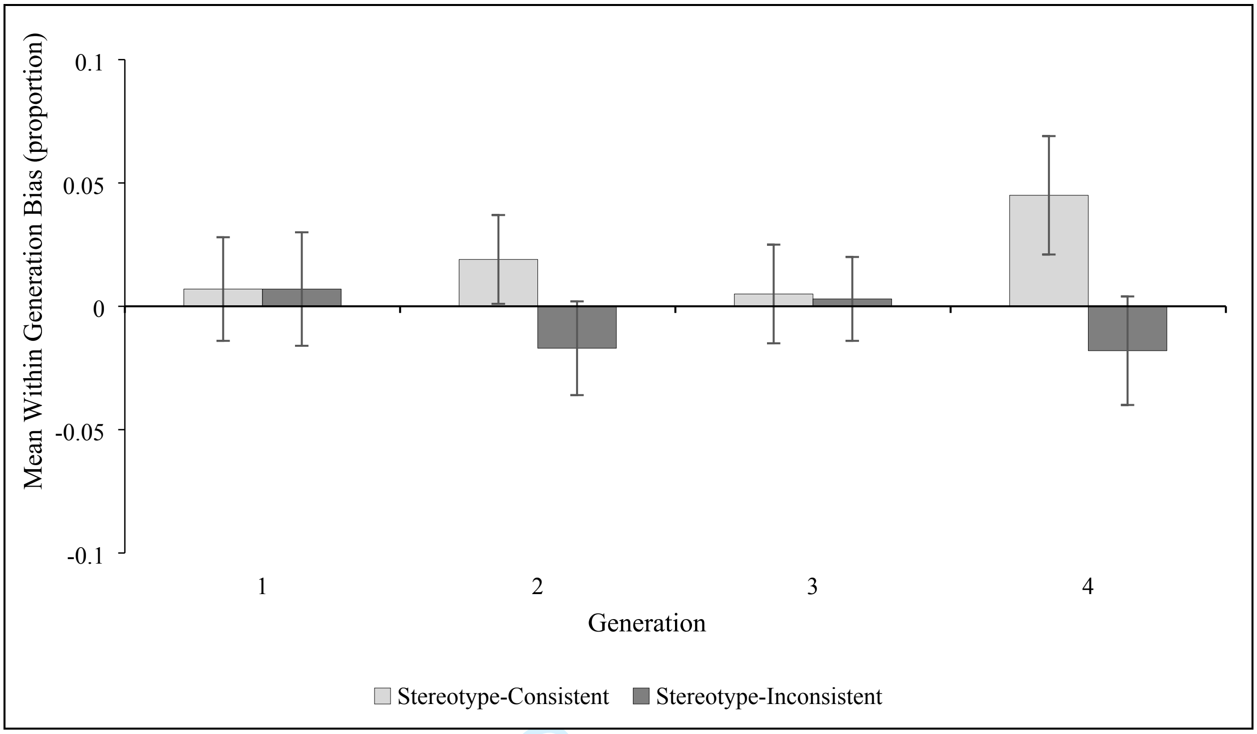


Figure 7. Expt. 1-3 within Generation bias by Experiment and Attribute Type. Bias scores were calculated by subtracting the proportion of stereotype-consistent attribute pairings from the training phase from the proportion of stereotype-consistent attributes pairings from the test phase. Error bars represent 95% confidence intervals.

Discussion – Cross-Experiment Analysis (Expt. 1-3)

The aim of the cross-experiment analysis was to establish whether the extent of stereotype maintenance differs dependent on target gender and social context and to further explore the mechanism underlying the effects. Supporting the trends from visual inspection of the means from Expt. 1-3, the cross-experiment analysis indicates there was stronger stereotype maintenance in the masculine context than in the neutral context and stronger stereotype maintenance for male targets than female targets. As information passed down the chains it became increasingly learnable, with significant increases in the proportion of stereotype-consistent and stereotype-inconsistent correct responses. Even though there was no evidence of a memory bias towards stereotype-consistent information at the start of the chains – in either correct recall or memory intrusions – a memory bias for stereotype-consistent information emerged by the end of the chains. There was evidence that the strength of the stereotype-consistent memory bias changed across the generations, with a larger bias towards stereotype-consistent target-attribute pairings in the final generation than in the first generation.

The cross-experiment analysis suggests social information evolves to become increasingly gender stereotype-consistent over time and this process is mediated by both target gender and a gendered social context (i.e., gendered occupations). Our findings are consistent with recent evidence that gender schema are relative social constructs, which vary in their characteristics dependent on the social context in which they are encountered (Martin, 2023). For example, Martin found that gender neutral personality attributes were perceived to be more feminine when they appeared alongside masculine attributes, but were perceived to be more masculine when they appeared alongside feminine attributes. Similarly, gender-neutral job candidates were perceived to be less masculine when they were preceded by a masculine candidate, but were perceived to be less feminine when they were preceded by a feminine candidate. Martin's findings suggest that gender schema are not objective and fixed, but are instead subjective and malleable dependent on the context in which they appear (Martin, 2023). This begs the question of how the gender diversity of a social context might influence stereotype maintenance; for example, would one expect gender stereotypes to spontaneously re-emerge in a context where all targets were women or where all targets were men? We explored the influence of social context on stereotype maintenance in Expt. 4.

Expt. 4 Single Gender Target Context

The aim of Experiment 4 was to examine whether the spontaneous re-emergence of gender stereotypes is influenced by the gender diversity of the social context in which they appear (i.e., single-sex context vs. mixed-sex context). Recent evidence suggests the effects of gender schema are dependent on the context in which stimuli appear, with targets perceived to be more feminine/masculine when they appear in a context with other stimuli that are more masculine/feminine respectively (Martin, 2023). There is also evidence that the influence of gender schema is diminished when faces appear in a single gender context (e.g., only women's faces) relative to when they appear in a mixed gender context (e.g., both women's and men' faces; Macrae & Cloutier, 2009; Martin, et al., 2024). If gender schema are not objective and fixed, but are instead subjective and malleable dependent on the context in which they appear (Macrae & Cloutier, 2009; Martin, 2023; Martin, et al., 2024), it seems likely that this would influence the re-emergence of stereotypes via cumulative cultural evolution.

In Expt. 4 we investigated whether the contextual relationship between social targets influenced the maintenance of gender stereotypes via cultural evolution. Specifically, we examined whether gender stereotypes would be less likely to re-emerge in an intragroup context when all targets were from a single gender category, relative to an intergroup context when targets are from two gender categories (as in the previous experiments). The intergroup context was identical to that used in Expt. 3; the intragroup context perceivers were asked to learn about 16 targets, all of whom were female *or* all of whom were male (as opposed to 8 females *and* 8 males in Expt. 1-3). We chose to use the masculine job context from Expt. 3, as this produced the strongest evidence of stereotype maintenance. If encountering social targets in an intragroup context diminishes the influence of categorical person perception and stereotypes (e.g., Macrae & Cloutier, 2009; Martin, 2023; Martin, et al., 2024), we would expect there to be less evidence of gender stereotype maintenance in the single-sex condition than in the mixed-sex condition. Specifically, we predicted that by the end of the chains there would be significantly more stereotype consistent target-attribute pairings in the mixed-sex condition than in the single-sex condition and significantly more stereotype inconsistent target-attribute pairings in the single-sex condition than in the mixed-sex condition.

Expt. 4 Methods

Participants. The participants were 256 18-30-year-olds (128 female and 128 female)[7], half of whom were students recruited from the University of Aberdeen, who were granted course credit in exchange for taking part in the experiment, and half of whom were recruited online via Prolific Academic (www.prolific.ac), who completed the experiment remotely via the online testing platform *Gorilla* (www.gorilla.sc; Anwyl-Irvine, Massonnié, Flitton, Kirkham & Evershed, 2018), and were compensated around UK£6/US$7 for their time

Materials. The materials used were similar to those of Experiment 3, with the following exceptions. In the single-sex target condition, rather than participants seeing sixteen targets, half of whom were female and half of whom were male, participants within a chain saw sixteen targets that were either all female or all male. To this end, we created two new sets of target face images; one set of all female targets comprising the previous eight female target faces plus an additional eight female target faces, and one set of all male targets comprising the previous eight male target faces plus an additional eight male target faces. As before, the additional target faces were taken from the Chicago Face Database (Ma et al., 2015) and were subject to the same selection criteria. The mixed-sex target condition was identical to Expt. 3 with one exception; to exclude the possibility that the effects from Expt. 1-3 were driven by very specific target faces, we replaced the target images from Expt. 1-3 with the additional face images used to create the single-sex target sets.

[7] As in Expt. 3, in Expt. 4 we recruited an equal number of female and male participants to minimise the possibility that the effects were being driven by participant sex or gender. Participants were tested in within participant gender chains; in half of the chains the targets were all female and in half the chains the targets were all male. While underpowered, preliminary exploratory analyses revealed no significant main effects or interactions involving Chain Sex (all Fs < 3.44, all ps $> .68$).

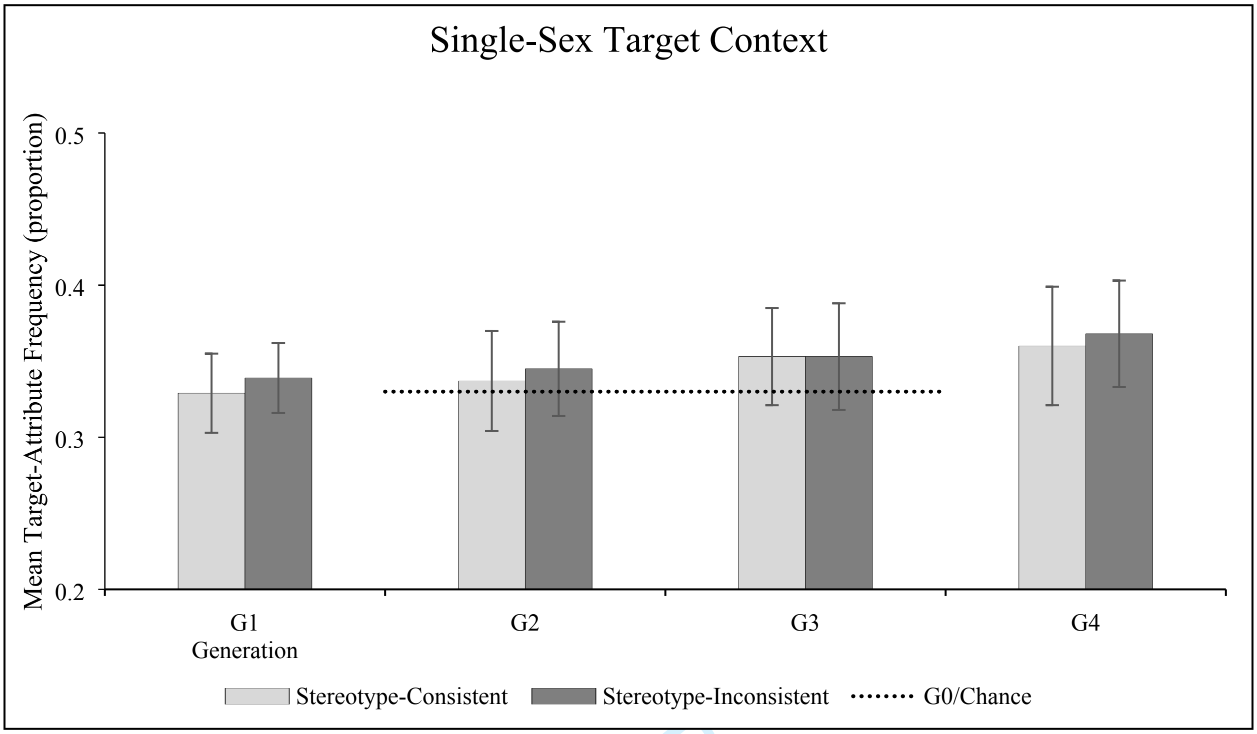
Procedure. For participants tested in the lab, the procedure was the same as that of Experiment 3. For participants tested online, we had no control over the physical environment in which they were tested. Participants were only able to complete the experiment on a laptop or desktop PC. The size and proportion of the face images were identical to Expt. 3 when presented on a 28-inch monitor with 1280 x 1024 screen resolution; the Gorilla testing platform standardizes the size ratio of the images across different screen sizes by asking participants to calibrate their screen size using a credit card.

Dependent Measures and Analyses. The dependent measures and analyses were identical to those used in Expt. 1-3, with the following exceptions; we added the between-subjects factor of Target Sex Diversity and removed the within-subjects factor of Target Sex (because the latter was a between-subjects variable in the single-sex target condition and a within-subjects variable in the mixed-sex target condition).

Expt. 4 – Results

We analysed the data for the mean proportional frequency with which targets were paired with gender stereotyped attributes at the start and end of the chains using a 2(Target Sex Context: single-sex targets vs. mixed-sex targets) X 2(Generation: G1 vs. G4) X 2(Attribute Type: stereotype-consistent vs. stereotype-inconsistent) mixed ANOVA with Target Sex Context as a between-subjects factor (see Supplementary Table 6 for all cell values and Supplementary Table 7 for cell values including Target Sex). The analysis revealed significant main effects of Generation [$F(1, 62) = 8.15$, $p = .006$; $\eta^2 = .116$] and Attribute Type [$F(1, 62) = 5.06$, $p = .028$; $\eta^2 = .075$], and significant interactions of Target Sex Context X Attribute Type [$F(1, 62) = 7.93$, $p = .007$; $\eta^2 = .113$] and Generation X Attribute Type [$F(1, 62) = 4.88$, $p = .031$; $\eta^2 = .073$]. However, all these effects were subsumed by the predicted significant interaction of Target Sex Context X Generation X Attribute Type [$F(1, 62) = 4.76$, $p = .033$;

$\eta^2 = .071$], which is explored below. There was no evidence of any other significant main effects or interactions (all *F*s < 2.10, all *p*s > .152).



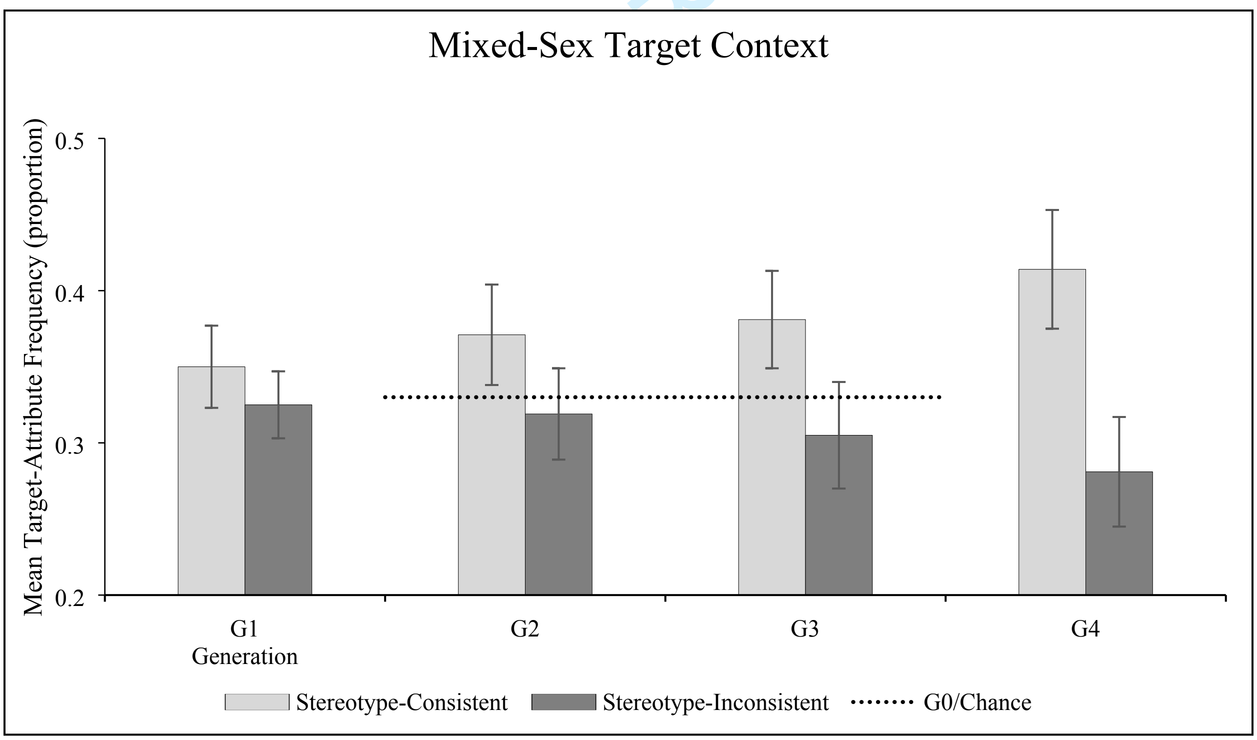


Figure 8. Expt. 4 frequency of target-attribute pairings by Generation and Attribute Type for the Single-Sex Context (top panel) and Mixed-Sex Context (bottom panel). Error bars represent 95% confidence intervals.

We explored the Target Sex Context X Generation X Attribute Type interaction by examining the pairwise comparisons for stereotype-consistent and stereotype-inconsistent target-attribute pairings within and between Generation 1 and Generation 4 for the mixed-sex target and single-sex target conditions respectively (Figure 8). The single-sex target condition diverged from Expt. 3, with no difference in stereotype consistent target-attribute pairings between G1 (*M prop* = .33; 95% CI [.30, .36]) and G4 (*M prop* = .36; 95% CI [.32, .40]; *M diff* = .03 (95% CI [-.01, .07]; $p = .163$; $d = .23$, 95% CI [-.13, .58]; $r = .163$, $p = .186$), and no difference in stereotype inconsistent target-attribute pairings between G1 (*M prop* = .34; 95% CI [.32, .36]) and G4 (*M prop* = .37; 95% CI [.33, .40]; *M diff* = .03 (95% CI [-.007, .067]; $p = .106$; $d = .27$, 95% CI [-.08, .63]; $r = .320$, $p = .037$). Whereas the mixed-sex target condition replicated the findings from Expt. 3, with a significant increase in stereotype consistent target-attribute pairings between G1 (*M prop* = .35; 95% CI [.32, .38]) and G4 (*M prop* = .41; 95% CI [.38, .45]; *M diff* = .06 (95% CI [.02, .11]; $p = .004$; $d = .59$, 95% CI [.21, .96]; $r = .173$, $p = .171$), and a significant decrease in stereotype inconsistent target-attribute pairings between G1 (*M prop* = .33; 95% CI [.30, .35]) and G4 (*M prop* = .28; 95% CI [.25, .32]; *M diff* = .04 (95% CI [.008, .08]; $p = .019$; $d = .46$, 95% CI [.09, .82]; $r = .264$, $p = .072$). Consequently, by the end of the chains, there was no difference in the frequency of stereotype-consistent target-attribute pairings and stereotype-inconsistent target attribute pairings in the single-sex condition (*M diff* = .01 (95% CI [-.06, .08]; $p = .799$; $d = .04$, 95% CI [-.31, .39]; $r = .632$, $p < .001$), whereas there were significantly more stereotype-consistent target-attribute pairings than stereotype -inconsistent target attribute pairings in the mixed-sex condition (*M diff* = .13 (95% CI [.07, .20]; $p < .001$; $d = .84$, 95% CI [.43, 1.24]; $r = -.458$, $p = .004$).

Discussion – Expt. 4

The aim of Experiment 4 was to examine whether the spontaneous re-emergence of gender stereotypes is influenced by the gender diversity of the social context in which they appear. The findings support and extend those from Expt. 1-3. As in Expt. 1-3, there was evidence information evolved as it passed along the chains, with significant changes in the frequency of stereotype-consistent and stereotype-inconsistent target-attribute pairings between the start and end of the chains. However, as predicted, this effect was modulated by the relative diversity of the social context with less evidence of gender stereotype maintenance in the single-sex condition than in the mixed-sex condition. Replicating the findings from Expt. 3, in a context containing both female and male targets gender stereotypes spontaneously re-emerged; however, in a single sex context, gender stereotypes did not re-emerge. These findings support recent evidence that gender schema are relative social constructs, which are less likely to bias social cognition in single gender social contexts (e.g., Macrae & Cloutier, 2009; Martin, 2023; Martin, et al., under review).

General Discussion

The aim of the current research was to establish whether societal stereotypes persist because stereotype-consistent memory bias incrementally shapes how social information evolves as it is repeatedly passed from person to person. Our findings support the possibility that, in the absence of communication bias, memory bias alone can cause social information to evolve to become increasingly stereotype consistent. While we found some evidence of the re-emergence of gender stereotypes in Expt. 1, the effects were stronger in a feminine stereotyped occupational context (Expt. 2) and in a masculine stereotyped occupational context (Expt. 3); conversely, the re-emergence of gender stereotypes was eliminated in a single gender context (Expt. 4). The current findings demonstrate that small amounts of cognitive bias, if widely shared, can cause existing gender stereotypes to be maintained through cultural evolution; however, they also suggest the maintenance of gender stereotypes is not the inevitable consequence of social transmission. Together, these findings have implications for our understanding of the memory biases that shape our individual experience and their cumulative consequences for the fabric of our society.

In-line with gender schema theory (Bem,1981; Liben & Bigler, 2017; Starr & Zurbriggen, 2017), the re-emergence of gender stereotypes was driven by a bias towards recalling stereotype-consistent information. Irrespective of whether a one endorses stereotypes, schematic knowledge of stereotype content has the potential to influence the way one attends to, stores in memory, and recalls social information (Bem, 1981; Tirado et al., 2023). Often this results in a bias towards information that is consistent with one's existing stereotype knowledge (Fyock & Stangor, 1994; Macrae & Bodenhausen, 2000; Rojahn & Pettigrew, 1992; Stangor & McMillan, 1992). In the current context, this bias towards gender stereotype-consistent information accumulated as information was repeatedly passed through the memories of the people in our chains. However, supporting the recent assertion that gender schema are relative social constructs (Martin, 2023), we also found that the extent to which gender stereotypes re-emerged was influenced by the social context in which targets appeared. Just as Martin found that perceptions of femininity/masculinity were dependent on the gendered context in which a stimulus appeared, so we found that the re-emergence of feminine/masculine stereotypes was dependent on the gendered context in which our targets appeared. Thus, at a macro-level the likelihood that gender stereotypes re-emerge via cultural evolution is ultimately dependent at a micro-level on the likelihood that gender stereotypes bias the memories of individual people.

The re-emergence of gender stereotypes was driven by a bias towards recalling stereotype-consistent information irrespective of the accuracy of these recollections. The cross-experiment analysis revealed that even though there was no evidence of a memory bias towards stereotype-consistent information at the start of the chains (in either correct recall or memory intrusions), a memory bias for stereotype-consistent information emerged by the end of the chains. Stereotypes re-emerged because people had a bias towards making stereotype-consistent intrusions (Bellezza & Bower, 1981; Lenton et al., 2001); indeed, it is unclear how the proportion of stereotype-consistent target-attribute pairings would have increased without a bias for stereotype-consistent intrusions. In reality, the nature of the current experiments is not well suited for determining whether people are better at accurately recalling stereotype-consistent information (Fyock & Stangor, 1994). This is because as the proportion of stereotype-consistent responses increases, so does the likelihood of higher accuracy for stereotype-consistent target-attribute pairings by chance. This issue is exacerbated as information passes down the chains, as a bias towards making more stereotype-consistent responses in the previous generation increases the proportion of stereotype-consistent target attribute pairings in the training phase of the current generation, which then further increases the likelihood of higher accuracy for stereotype-consistent target-attribute pairings by chance. Thus, the fact that the learning environment became increasingly stereotype-consistent might explain why people showed an increase in stereotype-consistent correct responses over time. However, irrespective of whether people are better at correctly recalling stereotype-consistent information, a bias towards stereotype-consistent memory intrusions can explain the re-emergence of stereotypes in the social environment.

The findings of the cross-experiment analysis suggest the strength of the stereotype-consistent memory bias increased as the social environment became increasingly stereotype-consistent. Participants at the end of the chains produced even more additional stereotype-consistent target attribute pairings than did participants at the start of the chains. As outlined above, at Generation 1, when the social environment did not contain gender stereotype-consistent bias, it appeared spontaneously through people's small bias towards stereotype-consistent memory intrusions. However, by Generation 4, when the social environment did contain some gender stereotype-consistent bias, the size of people's bias towards stereotype-consistent memory intrusions increased. Thus, the re-emergence of gender stereotypes at the end of the chains is not simply the accumulation of a uniform bias towards gender stereotype-consistent memory. It is an interactive process between the information in the environment and the way this information is cognitively represented. It is as though the stereotypes have been bootstrapped into re-emergence and that once they re-emerge, they exert an even greater influence on people's cognitive representations and therefore the evolving environment.

While the cumulative effects of a small but widely shared bias towards gender stereotype-consistent memory might have negative consequences for society, the bias itself and the fact that it is widely shared might have adaptive consequences for individuals (Hutchison & Martin, 2015). People are only able to attend to, store, and accurately recall from memory a tiny proportion of the information that encounter in everyday life. This means there are many circumstances in which people are likely to suffer memory lapses when recalling information about other people. If these memory lapses were independent of gender stereotypes, then, as in our experiment, sometimes they would be accurate and sometimes they would not. However, in a world where women and men are more likely to identify themselves with gender stereotype-consistent attributes (Donnelly & Twenge, 2017), a bias towards gender stereotype-consistent intrusions increases the likelihood that one's memory lapses will be accurate. In addition, if a bias towards gender stereotype-consistent memory intrusions is widely shared across the population, it increases the likelihood that people will share common ground with one another. Such stereotype-consistent common ground, even if based on error, can reduce uncertainty (Hogg & Reid, 2006), simplify communication (Stangor and Schaller, 2000), facilitate social interaction (Ruscher, 1998), and provide a foundation for a more structured and cohesive society (Bar-Tal, 2000).

Our findings suggest that even in the absence of social and communicative biases, bias for stereotype-consistent memory causes information to evolve to become increasing stereotype-consistent. Previous research suggests that people possess social and communicative biases that make them more likely to communicate stereotype-consistent information and that the effects of these biases accumulate over time (for a review see Kashima et al., 2007). It is possible that effects attributed to social or communicative bias might instead be driven by cognitive bias, as information must be recalled from memory before it can be communicated. However, in reality, it is likely that people possess independent cognitive, social, and communicative stereotype-consistent biases for stereotype-consistent information all of which contribute to the re-emergence of existing stereotypes. By adapting the transmission method we use here, it would be relatively straightforward for future research to disambiguate the distinct and combined contributions of these different forms of bias to the re-emergence and maintenance of societal stereotypes (for analogous work in language evolution see Kirby, Tamariz, Cornish, & Smith, 2015; Ota, San Jose, & Smith, 2021).

The current findings highlight the importance of considering the cumulative effects of psychological bias, even when the effects of this bias are apparently very small at an individual level (Mesoudi, 2011). At Generation 1, there was no evidence of a significant memory bias towards stereotype-consistent information relative to stereotype-inconsistent information in any of the four experiments. Yet in Expt. 1-3, the modest levels of stereotype-consistent bias we see at each generation accumulated into much more substantial bias through repeated social transmission. In Expt. 4, the single gendered context appeared to reduce the level of stereotype-consistent bias (either within or across individuals), thereby prevented the spontaneous re-emergence of gender stereotypes. At the current time there is a substantial gap in our understanding of how social bias influences society. On the one hand, there is a large and growing experimental literature documenting lab-based social cognitive bias. On the other hand, there is abundant evidence of real-world social bias in people's lived experience of individual and structural prejudice and discrimination. If many people share cognitive biases in the way they attend to, store, and recall social information (which we know they do), and if people repeatedly receive and transmit social information to others (which we know they do), then the likely consequence is directional cumulative cultural evolution (Hutchison & Martin, 2015; Martin et al., 2017). By examining the cumulative effects of social bias on culture, either through experimental work, through modelling, or through large-scale observational research, we might better understand how social cognitive bias continually shapes our society.

Interpretation and extrapolation of the findings from the present research are limited by the relative homogeneity of the social targets, workplace setting, and participant samples in our experiments. All targets in our experiments were White younger adults who worked in the an office environment; thus, gender was the only present primary category of social cognition (i.e., gender, race, and age; Fiske and Neuberg, 1990; Martin et al., 2015). Similarly, most participants were young, White women (although we did have equal numbers of men and women in Expts. 3-4). In reality, circumstances in which gender is the only relevant social category are rare. In typical real-world workplaces, for example, we might expect to see an array of ages, genders, and ethnicities, along with other, less primary categories such as job role, socio-economic status, religion, marital status, and sexual orientation. It would be worthwhile, then, to examine how stereotypes evolve when more than one such category is present – for example, in offices comprised of male and female, Black and White, younger and older adult co-workers, or among people who are employed in blue- and white-collar jobs, who are unemployed, or who are home caregivers (Tirado et al., 2023).

Conclusion

Across four experiments, we aimed to establish whether societal gender stereotypes re-emerge and persist because stereotype-consistent memory bias shapes how social information evolves as it is repeatedly passed from person to person. The current findings demonstrate that, despite apparently changing societal attitudes and beliefs about gender, even small amounts of cognitive bias, if widely shared, might cause existing stereotype content to be maintained through cumulative cultural evolution. Thus, even if the external social environment were free from stereotypes, people's gender schema mean that their cognitive representation of this environment would likely drive the spontaneous re-emergence of stereotypes as information is repeatedly socially transmitted. Yet, at the same time, our findings also suggest that this is not necessarily an inevitable consequence, and the way the social environment is structured might reduce the likelihood that stereotypes persist.

References

Abele, A. E., & Hauke, N. (2020). Comparing the facets of the big two in global evaluation of self versus other people. European Journal of Social Psychology, 50(5), 969-982.

Anwyl-Irvine, A., Massonnié, J., Flitton, A., Kirkham, N., & Evershed, J. (2018). Gorillas in our midst: Gorilla. sc. Behavior Research Methods,

Atli, A. (2017). High School Students' Gender Role Perceptions Regarding Various Professions. International Journal of Progressive Education, 13(3), 6-15.

Bar-Tal, D. (2000). Shared beliefs in a society: Social psychological analysis. Sage Publications.

Begeny, C. T., Ryan, M. K., Moss-Racusin, C. A., & Ravetz, G. (2020). In some professions, women have become well represented, yet gender bias persists—Perpetuated by those who think it is not happening. Science Advances, 6(26), eaba7814.

Bellezza, F. S., & Bower, G. H. (1981). Person stereotypes and memory for people. Journal of Personality and Social Psychology, 41(5), 856.

Bem, S. L. (1981). Bem sex-role inventory: Professional manual. Consulting Psychologists Press.

Bodenhausen, G. V., & Macrae, C. N. (2013). Stereotype activation and inhibition. Stereotype activation and inhibition (pp. 9-60). Psychology Press.

Boyd, R., & Richerson, P. J. (1988). Culture and the evolutionary process. University of Chicago press.

Clark, A. E., & Kashima, Y. (2007). Stereotypes help people connect with others in the community: A situated functional analysis of the stereotype consistency bias in communication. Journal of Personality and Social Psychology, 93(6), 1028.

Cundiff, J. L., & Vescio, T. K. (2016). Gender stereotypes influence how people explain gender disparities in the workplace. Sex Roles, 75(3), 126-138.

Devine, P. G. (1989). Stereotypes and prejudice: Their automatic and controlled components. Journal of Personality and Social Psychology, 56, 5-18. 10.1037/0022-3514.56.1.5

Donnelly, K., & Twenge, J. M. (2017). Masculine and feminine traits on the Bem Sex-Role Inventory, 1993–2012: A cross-temporal meta-analysis. Sex Roles, 76(9), 556-565.

Eagly, A. H., Nater, C., Miller, D. I., Kaufmann, M., & Sczesny, S. (2020). Gender stereotypes have changed: A cross-temporal meta-analysis of US public opinion polls from 1946 to 2018. American Psychologist, 75(3), 301-315.

Eagly, A. H., Wood, W., & Diekman, A. B. (2000). Social role theory of sex differences and similarities: A current appraisal. The Developmental Social Psychology of Gender, 12, 123-174.

Echterhoff, G., Higgins, E. T., & Groll, S. (2005). Audience-tuning effects on memory: the role of shared reality. Journal of Personality and Social Psychology, 89(3), 257-276.

Ellemers, N. (2018). Gender stereotypes. Annual Review of Psychology, 69, 275-298.

Faul, F., Erdfelder, E., Lang, A., & Buchner, A. (2007). G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behavior Research Methods, 39(2), 175-191.

Fiske, S. T., & Neuberg, S. L. (1990). A continuum model of impression formation from category based to individuating processes: Influences of information and motivation on attention and interpretation. In M. P. Zanna (Ed.), Advances in experimental social psychology (pp. 1-74). Academic Press.

Fyock, J., & Stangor, C. (1994). The role of memory biases in stereotype maintenance. British Journal of Social Psychology, 33(3), 331-343.

Greenwald, A. G., & Banaji, M. R. (1995). Implicit social cognition: attitudes, self-esteem, and stereotypes. Psychological Review, 102(1), 4-27.

Haines, E. L., Deaux, K., & Lofaro, N. (2016). The times they are a-changing… or are they not? A comparison of gender stereotypes, 1983–2014. Psychology of Women Quarterly, 40(3), 353-363.

Hogg, M. A., & Reid, S. A. (2006). Social identity, self-categorization, and the communication of group norms. Communication Theory, 16(1), 7-30.

Hutchison, J., Cunningham, S. J., Slessor, G., Urquhart, J., Smith, K., & Martin, D. (2018). Context and perceptual salience influence the formation of novel stereotypes via cumulative cultural evolution. Cognitive Science, 42, 186-212.

Hutchison, J., & Martin, D. (2015). The evolution of stereotypes. Evolutionary perspectives on social psychology (pp. 291-301). Springer.

Kashima, Y., Klein, O., & Clark, A. E. (2007). Grounding: Sharing Information in Social Interaction. In K. Fiedler (Ed.), Social communication (pp. 27-77). Psychology Press.

Kashima, Y. (2000). Maintaining cultural stereotypes in the serial reproduction of narratives. Personality and Social Psychology Bulletin, 26(5), 594-604.

Kashima, Y., Lyons, A., & Clark, A. (2013). The maintenance of cultural stereotypes in the conversational retelling of narratives. Asian Journal of Social Psychology, 16(1), 60-70.

Kirby, S., Cornish, H., & Smith, K. (2008). Cumulative cultural evolution in the laboratory: An experimental approach to the origins of structure in human language. Proceedings of the National Academy of Sciences of the United States of America, 105(31), 10681-10686. 10.1073/pnas.0707835105

Kirby, S., Tamariz, M., Cornish, H., & Smith, K. (2015). Compression and communication in the cultural evolution of linguistic structure. Cognition, 141, 87-102.

Lenton, A. P., Blair, I. V., & Hastie, R. (2001). Illusions of gender: Stereotypes evoke false memories. Journal of Experimental Social Psychology, 37(1), 3-14.

Liben, L. S., & Bigler, R. S. (2017). Understanding and undermining the development of gender dichotomies: The legacy of Sandra Lipsitz Bem. Sex Roles, 76, 544-555.

Lueptow, L. B., Garovich-Szabo, L., & Lueptow, M. B. (2001). Social change and the persistence of sex typing: 1974–1997. Social Forces, 80(1), 1-36.

Lyons, A., & Kashima, Y. (2003). How are stereotypes maintained through communication? The influence of stereotype sharedness. Journal of Personality and Social Psychology, 85(6), 989-1005. 10.1037/0022-3514.85.6.989

Ma, D. S., Correll, J., & Wittenbrink, B. (2015). The Chicago face database: A free stimulus set of faces and norming data. Behavior Research Methods, 47(4), 1122-1135.

Macrae, C. N., & Bodenhausen, G. V. (2000). Social cognition: Thinking categorically about others. Annual Review of Psychology, 51(1), 93-120.

Madon, S., Guyll, M., Aboufadel, K., Montiel, E., Smith, A., Palumbo, P., & Jussim, L. (2001). Ethnic and national stereotypes: The Princeton trilogy revisited and revised. Personality and Social Psychology Bulletin, 27(8), 996-1010.

Martin, D., Hutchison, J., Slessor, G., Urquhart, J., Cunningham, S. J., & Smith, K. (2014). The spontaneous formation of stereotypes via cumulative cultural evolution. *Psychological Science, 25*(9), 1777-1786. 10.1177/0956797614541129

Martin, D., Cunningham, S. J., Hutchison, J., Slessor, G., & Smith, K. (2017). How societal stereotypes might form and evolve via cumulative cultural evolution. *Social and Personality Psychology Compass, 11*(9), e12338.

Martin, D., Hutchison, J., Konopka, A., Dallimore, C. J., Slessor, G., & Swainson, R. (2024). Intergroup processes and the happy face advantage: How social categories influence emotion categorization. *Journal of Personality and Social Psychology, 126(3),* 390-412.

Martin, D., Swainson, R., Slessor, G., Hutchison, J., Marosi, D., & Cunningham, S. J. (2015). The simultaneous extraction of multiple social categories from unfamiliar faces. *Journal of Experimental Social Psychology, 60*, 51-58.

Matheus, C. C., & Quinn, E. (2017). Gender based occupational stereotypes: New behaviors, old attitudes. Paper presented at the 2017 IEEE Women in Engineering (WIE) Forum USA East, 1-6.

Mesoudi, A. (2011). Cultural evolution: How Darwinian theory can explain human culture and synthesize the social sciences. University of Chicago Press.

Nosek, B. A., Smyth, F. L., Sriram, N., Lindner, N. M., Devos, T., Ayala, A., Bar-Anan, Y., Bergh, R., Cai, H., Gonsalkorale, K., Kesebir, S., Maliszewski, N., Neto, F., Olli, E., Park, J., Schnabel, K., Shiomura, K., Tulbure, B. T., Wiers, R. W., . . . Greenwald, A. G. (2009). National differences in gender-science stereotypes predict national sex differences in science and math achievement. Proceedings of the National Academy of Sciences of the United States of America, 106(26), 10593-10597. 10.1073/pnas.0809921106

Ota, M., San José, A., & Smith, K. (2021). The emergence of word-internal repetition through iterated learning: Explaining the mismatch between learning biases and language design. Cognition, 210, 104585.

Rojahn, K., & Pettigrew, T. F. (1992). Memory for schema-relevant information: A meta-analytic resolution. British Journal of Social Psychology, 31(2), 81-109.

Ruscher, J. B. (1998). Prejudice and stereotyping in everyday communication. Advances in experimental social psychology (pp. 241-307). Elsevier.

Sargent, P. (2000). Real men or real teachers? Contradictions in the lives of men elementary teachers. Men and Masculinities, 2(4), 410-433.

Sebastián-Tirado, A., Félix-Esbrí, S., Forn, C., & Sanchis-Segura, C. (2023). Gender Stereotypes Selectively Affect the Remembering of Highly Valued Professions. Sex Roles, 88(7-8), 326-347.

Simpson, R. (2004). Masculinity at work: The experiences of men in female dominated occupations. Work, Employment and Society, 18(2), 349-368.

Sperber, D. (1996). Explaining culture. Blackwell Publishers.

Stangor, C., & McMillan, D. (1992). Memory for expectancy-congruent and expectancy-incongruent information: A review of the social and social developmental literatures. Psychological Bulletin, 111(1), 42-61.

Stangor, C., & Schaller, M. (2000). Stereotypes as individual and collective representations. In C. Stangor (Ed.), Stereotypes and prejudice: Essential readings. (pp. 64-82). Psychology Press.

Starr, C. R., & Zurbriggen, E. L. (2017). Sandra Bem's gender schema theory after 34 years: A review of its reach and impact. Sex Roles, 76, 566-578.

Washington, G. (1931). The writings of George Washington from the original manuscript sources, 1745-1799. US Government Printing Office.

Wigboldus, D. H., Semin, G. R., & Spears, R. (2000). How do we communicate stereotypes? Linguistic bases and inferential consequences. Journal of Personality and Social Psychology, 78(1), 5-18.

Wigboldus, D. H., Sherman, J. W., Franzese, H. L., & Knippenberg, A. v. (2004). Capacity and comprehension: Spontaneous stereotyping under cognitive load. Social Cognition, 22(3), 292-309.

Supplementary Table A. Target face image norms. Target image name refers to the original from the Chicago Face Database (Ma et al., 2015). Age is in years. All other measures are mean ratings on a 5-point Likert scale (1 = Least – 5 = Most)

Expt.	Target Image Name	Target Sex	Feminine	Masculine	Attractive	Trustworthy	Unusual	Age
1 to 4	WF-008	F	4.00	2.27	2.73	3.88	1.68	30.21
1 to 4	WF-215	F	4.25	2.24	3.28	2.96	2.24	30.64
1 to 4	WF-232	F	5.03	1.28	4.21	3.82	1.72	25.28
1 to 4	WF-243	F	5.28	1.56	4.36	4.00	2.83	26.48
1 to 4	WF-023	F	3.96	2.37	2.74	3.57	2.11	32.74
1 to 4	WF-213	F	4.14	2.10	3.24	3.00	1.86	25.66
1 to 4	WF-011	F	5.14	1.70	4.13	3.70	1.83	24.04
1 to 4	WF-207	F	5.25	1.46	4.46	3.96	2.04	24.88
1 to 4	WM-016	M	1.41	5.21	3.26	3.08	1.95	30.40
1 to 4	WM-238	M	1.50	4.59	3.27	3.00	2.14	24.82
1 to 4	WM-009	M	2.03	4.73	4.08	3.51	2.23	23.70
1 to 4	WM-029	M	1.61	5.21	4.59	3.70	2.05	28.59
1 to 4	WM-218	M	1.36	4.39	2.86	3.11	1.89	25.82
1 to 4	WM-203	M	1.33	4.87	3.50	3.54	1.54	29.25
1 to 4	WM-212	M	1.81	4.96	3.70	3.48	2.52	30.11
1 to 4	WM-004	M	1.80	4.85	4.66	3.56	1.93	25.82
4	WF-001	F	3.86	3.01	3.11	3.30	2.80	24.95
4	WF-027	F	5.31	1.51	4.69	3.96	2.07	21.29
4	WF-009	F	4.07	2.44	3.19	3.44	2.69	23.31
4	WF-229	F	3.92	2.29	2.68	3.20	3.48	22.92
4	WF-019	F	3.98	2.81	2.93	3.11	2.61	28.35
4	WF-247	F	4.83	1.58	3.25	3.61	1.67	26.42
4	WF-231	F	5.31	1.45	3.86	3.34	1.72	26.31
4	WF-024	F	5.34	1.74	4.76	3.73	2.44	23.94
4	WM-040	M	1.81	4.66	3.05	3.42	2.33	25.12
4	WM-214	M	2.25	3.67	3.12	3.72	2.24	25.12
4	WM-250	M	1.56	4.68	4.12	3.36	2.24	21.12
4	WM-207	M	1.33	5.21	4.13	3.33	2.46	26.58
4	WM-219	M	1.79	4.58	2.96	3.25	1.63	22.17
4	WM-210	M	1.68	3.48	2.96	3.32	2.07	21.50
4	WM-213	M	1.92	4.16	3.71	3.56	2.32	26.36
4	WM-257	M	1.81	4.35	3.74	3.92	2.67	30.04

Supplementary Table B. Results of independent samples t-tests examining differences in mean ratings between Female and Male target faces. Age is in years; all other measures are Likert responses (1 = Most – 5 = Least).

	Female Targets		Male Targets		*t(30)*	*p*	Cohen's d
	M	*SD*	*M*	*SD*			
Femininity	4.6	0.6	1.6	0.2	17.54	< .001	6.20
Masculinity	1.9	0.4	4.9	0.3	14.58	< .001	5.15
Attractiveness	3.6	0.7	3.8	0.6	.053	.958	.02
Trustworthiness	3.6	0.4	3.4	0.3	.980	.334	.35
Unusualness	2	0.4	2	0.3	.693	.494	.25
Age	27.5	3.2	27.3	2.6	.041	.968	.01

Supplementary Table C. Mean gender stereotype ratings for attributes from pilot testing (1 = feminine – 5 = masculine). During pilot testing, we asked a sample of 151 undergraduate students, who did not take part in the Experiments proper, to rate the relative gender stereotype of each attribute. The instructions given to participants can be seen overleaf.

Attribute	*N*	*M*	*SD*	Attribute Type
caring	151	1.97	.82	Feminine
sympathetic	151	2.06	.76	Feminine
compassionate	150	2.06	.77	Feminine
affectionate	150	2.09	.86	Feminine
soft spoken	151	2.12	.85	Feminine
gentle	151	2.20	.86	Feminine
innocent	151	2.25	.81	Feminine
warm	151	2.26	1.00	Feminine
kind	150	2.31	.85	Feminine
understanding	151	2.35	.76	Feminine
encouraging	151	2.40	.80	Feminine
gullible	151	2.46	.91	Feminine
polite	151	2.50	.84	Feminine
shy	150	2.53	.76	Feminine
cheerful	150	2.55	.75	Feminine
loyal	151	2.56	.68	Feminine
jealous	151	2.60	.87	Neutral
anxious	151	2.61	.76	Neutral
tactful	150	2.65	.74	Neutral
bitter	150	2.68	.82	Neutral
committed	151	2.74	.78	Neutral
friendly	151	2.76	.66	Neutral
truthful	150	2.85	.73	Neutral
secretive	151	2.92	.90	Neutral
nasty	151	2.97	.93	Neutral
lonely	151	2.99	.81	Neutral
unpleasant	151	3.07	.99	Neutral
selfish	151	3.07	.66	Neutral
troublesome	151	3.11	.96	Neutral
bullying	151	3.13	.67	Neutral
thoughtless	150	3.17	.93	Neutral
offensive	151	3.23	1.03	Neutral
original	151	3.29	.85	Masculine
independent	151	3.31	.80	Masculine
strong willed	151	3.36	.96	Masculine
self-reliant	151	3.40	.97	Masculine
strong personality	151	3.40	.78	Masculine
analytical	151	3.40	.72	Masculine
hostile	151	3.45	.84	Masculine
self-sufficient	149	3.50	.84	Masculine
confident	151	3.56	.87	Masculine
ambitious	150	3.57	.84	Masculine
assertive	150	3.61	.89	Masculine
decisive	151	3.70	.98	Masculine
competitive	151	3.73	.86	Masculine
forceful	151	3.92	.80	Masculine
dominant	151	3.99	.86	Masculine
arrogant	150	4.05	.82	Masculine

Supplementary Table D. Results of paired samples t-tests examining differences in pilot test mean gender stereotype ratings for attributes grouped as feminine, masculine, and neutral (1 = feminine – 5 = masculine)

	N	*M*	*SD*	*t*	*p*	*d*
Feminine vs.	142	2.27	.44	19.58	< .001	1.64
Masculine	142	3.59	.45			
Feminine vs.	145	2.28	.48	14.95	< .001	1.24
Neutral	145	2.91	.28			
Masculine vs.	144	3.60	.44	17.47	< .001	1.46
Neutral	144	2.90	.28			

Instructions for pilot ratings of attributes:

In this survey we are interested in your knowledge of whether certain words are associated with the feminine and masculine stereotypes.

IMPORTANT - WE ARE NOT INTERESTED IN YOUR OWN PERSONAL BELIEFS

Please indicate how stereotypically feminine/masculine each of the words below are using the following scale:

1 = strongly associated with feminine stereotype

2 = more strongly associated with feminine than masculine stereotype

3 = not associated with either stereotype/equally associated with both stereotypes

4 = more strongly associated with masculine than feminine stereotype

5 = strongly associated with masculine stereotype

Supplementary Table E. Mean pilot ratings of how strongly associated personality attributes are with the fictitious Feminine and Masculine job titles. During pilot testing, we asked a sample of 180 undergraduate students (45 per job condition), who did not take part in the Experiments proper, to rate the relative gender stereotype of each attribute. The instructions given to participants can be seen overleaf.

Feminine Stereotype Jobs			Masculine Stereotyped Jobs		
Attribute	*M*	*SD*	Attribute	*M*	*SD*
friendly (n)	1.70	1.08	analytical (m)	1.99	1.23
caring (f)	1.74	1.08	committed (n)	2.15	1.17
committed (n)	1.75	1.07	ambitious (m)	2.16	1.08
encouraging (f)	1.78	1.23	confident (m)	2.19	1.07
polite (f)	1.81	1.03	self-reliant (m)	2.34	1.15
understanding (f)	1.81	1.10	selfish (n)	2.35	1.17
kind (f)	1.82	1.04	assertive (m)	2.36	1.01
compassionate (f)	1.86	1.06	independent (m)	2.39	1.17
sympathetic (f)	1.86	1.10	decisive (m)	2.41	1.19
warm (f)	1.94	1.08	strong personality (m)	2.50	.99
confident (m)	2.03	1.12	competitive (m)	2.53	1.07
loyal (f)	2.05	.99	strong-willed (m)	2.53	1.08
truthful (n)	2.14	1.14	truthful (n)	2.55	1.12
affectionate (f)	2.17	.94	encouraging (f)	2.58	.98
cheerful (f)	2.17	1.07	original (m)	2.59	1.06
gentle (f)	2.18	.93	tactful (n)	2.64	1.09
tactful (n)	2.29	1.24	loyal (f)	2.66	1.10
ambitious (m)	2.34	1.01	understanding (f)	2.70	1.14
decisive (m)	2.44	1.11	friendly (n)	2.71	1.02
self-reliant (m)	2.45	1.13	polite (f)	2.79	1.02
self-sufficient (m)	2.47	1.08	dominant (m)	2.81	1.14
original (m)	2.61	.89	caring (f)	2.89	1.14
analytical (m)	2.62	1.11	cheerful (f)	2.93	.96
strong-willed (m)	2.65	1.10	kind (f)	2.98	.97
independent (m)	2.68	1.17	sympathetic (f)	2.98	1.07
strong personality (m)	2.71	1.06	warm (f)	3.03	.98
assertive (m)	2.78	1.07	compassionate (f)	3.06	1.01
soft-spoken (f)	3.01	1.24	forceful (m)	3.06	1.17
innocent (f)	3.06	1.02	secretive (n)	3.18	1.17
competitive (m)	3.38	1.16	affectionate (f)	3.23	.97
dominant (m)	3.40	1.00	gentle (f)	3.23	.94
forceful (m)	3.84	1.08	innocent (f)	3.33	.98
gullible (f)	3.95	1.11	arrogant (m)	3.39	1.23
shy (f)	4.00	1.11	self-sufficient (m)	3.39	1.22
secretive (n)	4.01	1.16	soft-spoken (f)	3.58	1.05
anxious (n)	4.09	1.04	offensive (n)	3.60	1.37
arrogant (m)	4.09	1.14	anxious (n)	3.65	1.09
lonely (n)	4.12	1.14	shy (f)	3.65	1.16
unpleasant (n)	4.16	1.21	unpleasant (n)	3.65	1.21
jealous (n)	4.19	1.10	troublesome (n)	3.66	1.36
thoughtless (n)	4.21	1.21	hostile (m)	3.68	1.28
bitter (n)	4.22	1.22	lonely (n)	3.68	1.14
selfish (n)	4.23	1.23	bitter (n)	3.73	1.23
troublesome (n)	4.23	1.16	bullying (n)	3.74	1.38
hostile (m)	4.34	1.15	nasty (n)	3.76	1.29
offensive (n)	4.40	1.09	jealous (n)	3.80	1.21
bullying (n)	4.45	1.09	thoughtless (n)	3.91	1.29
nasty (n)	4.45	1.06	gullible (f)	4.00	1.23

Supplementary Table F. Results of paired samples t-tests examining whether each fictitious job title was more strongly associated with stereotypically feminine or masculine attributes (1 = most strongly associated – 5 = least strongly associated)

		Feminine Attributes		Masculine Attributes				
	N	*M*	*SD*	*M*	*SD*	*t*	*p*	Cohen's *d*
Mobility Analyst	37	3.07	.75	2.63	.67	2.44	.020	.40
Product Factors	44	3.12	.55	2.54	.50	5.02	< .001	.77
Social Resources	38	2.44	.69	2.92	.58	3.21	.001	.52
Community Outreach	39	2.21	.61	2.94	.49	7.33	< .001	1.17

Example of instructions for pilot ratings of attributes associated with fictitious job titles:

Which personality attributes do you associate with people who work in a "Mobility Analysts Team"?

In this survey we are interested in your opinion of how strongly associated personality attributes are with a member of a "Mobility Analysts Team" using the following scale:

1 = Definitely associated with "Mobility Analysts Team"

2 = Probably associated with "Mobility Analysts Team"

3 = Neutral

4 = Probably NOT associated with "Mobility Analysts Team"

5 = Definitely NOT associated with "Mobility Analysts Team"

Supplementary Table 1. Expt. 1 Mean proportion frequency of target-attribute pairings by Generation, Attribute Type, and Target Sex

		Female Targets			Male Targets		
		Stereotype Consistent	Stereotype Inconsistent	Neutral	Stereotype Consistent	Stereotype Inconsistent	Neutral
Generation 1	Mean	.280	.379	.341	.333	.337	.329
	SD	.084	.112	.085	.111	.128	.083
	N	16	16	16	16	16	16
Generation 2	Mean	.319	.361	.320	.387	.314	.299
	SD	.104	.124	.063	.106	.118	.107
	N	16	16	16	16	16	16
Generation 3	Mean	.306	.401	.293	.374	.340	.286
	SD	.130	.137	.077	.118	.116	.060
	N	16	16	16	16	16	16
Generation 4	Mean	.358	.376	.266	.411	.335	.254
	SD	.122	.102	.098	.151	.141	.072
	N	16	16	16	16	16	16

Supplementary Table 2. Expt. 2 Mean proportion frequency of target-attribute pairings by Generation, Attribute Type, and Target Sex

		Female Targets			Male Targets		
		Stereotype Consistent	Stereotype Inconsistent	Neutral	Stereotype Consistent	Stereotype Inconsistent	Neutral
Generation 1	Mean	.346	.361	.293	.368	.323	.309
	SD	.064	.094	.068	.108	.106	.074
	N	16	16	16	16	16	16
Generation 2	Mean	.350	.332	.318	.383	.306	.311
	SD	.130	.091	.123	.124	.131	.090
	N	16	16	16	16	16	16
Generation 3	Mean	.365	.311	.324	.359	.289	.352
	SD	.132	.113	.097	.097	.096	.101
	N	16	16	16	16	16	16
Generation 4	Mean	.408	.299	.293	.406	.299	.294
	SD	.139	.118	.099	.117	.093	.110
	N	16	16	16	16	16	16

Supplementary Table 3. Expt. 3 Mean proportion frequency of target-attribute pairings by Generation, Attribute Type, and Target Sex

		Female Targets			Male Targets		
		Stereotype Consistent	Stereotype Inconsistent	Neutral	Stereotype Consistent	Stereotype Inconsistent	Neutral
Generation 1	Mean	.335	.319	.346	.378	.320	.302
	SD	.071	.063	.065	.129	.107	.066
	N	16	16	16	16	16	16
Generation 2	Mean	.335	.353	.312	.383	.272	.345
	SD	.059	.099	.084	.119	.097	.123
	N	16	16	16	16	16	16
Generation 3	Mean	.336	.363	.301	.448	.250	.302
	SD	.093	.112	.079	.127	.118	.092
	N	16	16	16	16	16	16
Generation 4	Mean	.375	.314	.311	.499	.221	.280
	SD	.126	.118	.091	.121	.100	.107
	N	16	16	16	16	16	16

Supplementary Table 4. Expt. 1-3 Mean proportion frequency of target-attribute pairings by Generation, Attribute Type, and Target Sex

		Female Targets			Male Targets		
		Stereotype Consistent	Stereotype Inconsistent	Neutral	Stereotype Consistent	Stereotype Inconsistent	Neutral
Generation 1	Mean	.320	.353	.327	.360	.327	.313
	SD	.078	.094	.076	.115	.112	.074
	N	48	48	48	48	48	48
Generation 2	Mean	.335	.349	.317	.384	.297	.319
	SD	.101	.104	.092	.114	.115	.107
	N	48	48	48	48	48	48
Generation 3	Mean	.336	.359	.306	.394	.293	.313
	SD	.120	.124	.084	.119	.114	.089
	N	48	48	48	48	48	48
Generation 4	Mean	.380	.330	.290	.439	.285	.276
	SD	.128	.115	.096	.135	.121	.097
	N	48	48	48	48	48	48

Supplementary Table 5. Cross-experiment ANOVA results. 3(Experiment Context: neutral vs. feminine vs. masculine) X 2(Generation: G1 vs. G4) X 2(Target Sex: female targets vs. male targets) X 2(Attribute Type: stereotype-consistent vs. stereotype-inconsistent) X 2(Accuracy: correct vs. intrusions).

Effect	Sum of Squares	*df*	*MSE*	*F*	*p*	η^2
Intercept	23.42	1	23.42	6107.00	<.001	.993
Experiment	.002	2	.001	.22	.805	.010
Error (Experiment)	.173	45	.004			
Generation	.017	1	.017	13.27	<.001	.228
Generation * Experiment	.010	2	.005	3.93	.027	.149
Error (Generation)	.056	45	.001			
Accuracy	3.252	1	3.252	386.27	<.001	.896
Accuracy * Experiment	.065	2	.032	3.85	.029	.146
Error (Accuracy)	.379	45	.008			
Attribute Type	.125	1	.125	9.45	.004	.174
Attribute Type * Experiment	.106	2	.053	4.01	.025	.151
Error (Attribute Type)	.597	45	.013			
Target Sex	.002	1	.002	1.72	.196	.037
Target Sex * Experiment	.004	2	.002	1.65	.204	.068
Error (Target Sex)	.059	45	.001			
Generation * Accuracy	.546	1	.546	88.87	<.001	.664
Generation * Accuracy * Experiment	.066	2	.033	5.38	.008	.193
Error (Generation*Accuracy)	.277	45	.006			
Generation * Attribute Type	.124	1	.124	18.60	<.001	.292
Generation * Attribute Type * Experiment	.006	2	.003	.45	.640	.020
Error (Generation*Attribute Type)	.301	45	.007			
Accuracy * Attribute Type	.015	1	.015	2.25	.140	.048
Accuracy * Attribute Type * Experiment	.021	2	.010	1.60	.213	.067
Error (Accuracy*Attribute Type)	.293	45	.007			
Generation * Accuracy * Attribute Type	.043	1	.043	5.84	.020	.115
Generation * Accuracy * Attribute Type * Experiment	.014	2	.007	.93	.401	.040
Error (Generation*Accuracy*Attribute Type)	.329	45	.007			
Generation * Target Sex	.000	1	.000	.00	.982	.000
Generation * Target Sex * Experiment	.000	2	.000	.17	.842	.008
Error (Generation*Target Sex)	.049	45	.001			
Accuracy * Target Sex	.031	1	.031	12.36	.001	.215
Accuracy * Target Sex * Experiment	.001	2	.000	.15	.865	.006
Error (Accuracy*Target Sex)	.112	45	.002			
Generation * Accuracy * Target Sex	.000	1	.000	.11	.743	.002
Generation * Accuracy * Target Sex * Experiment	.000	2	.000	.08	.925	.003
Error (Generation*Accuracy*Target Sex)	.123	45	.003			
Attribute Type * Target Sex	.086	1	.086	6.29	.016	.123
Attribute Type * Target Sex * Experiment	.021	2	.010	.76	.476	.032
Error (Attribute Type*Target Sex)	.612	45	.014			
Generation * Attribute Type * Target Sex	.004	1	.004	.49	.488	.011
Generation * Attribute Type * Target Sex * Experiment	.030	2	.015	1.71	.192	.071
Error (Generation*Attribute Type*Target Sex)	.394	45	.009			
Accuracy * Attribute Type * Target Sex	.021	1	.021	2.99	.091	.062
Accuracy * Attribute Type * Target Sex * Experiment	.011	2	.006	.83	.442	.036
Error (Accuracy*Attribute Type*Target Sex)	.310	45	.007			
Generation * Accuracy * Attribute Type * Target Sex	.000	1	.000	.04	.853	.001
Generation * Accuracy * Attribute Type * Target Sex * Experiment	.001	2	.000	.05	.952	.002
Error (Generation*Accuracy*Attribute Type*Target Sex)	.263	45	.006			

Supplementary Table 6. Expt. 6 Mean proportion frequency of target-attribute pairings by Target Sex Context, Generation, Target Sex, and Attribute Type

		Single Sex Targets			Mixed Sex Targets		
		Stereotype Consistent	Stereotype Inconsistent	Neutral	Stereotype Consistent	Stereotype Inconsistent	Neutral
Generation 1	Mean	.329	.339	.333	.350	.326	.320
	SD	.079	.060	.069	.072	.068	.070
	N	32	32	32	32	32	32
Generation 2	Mean	.337	.345	.319	.371	.319	.304
	SD	.103	.097	.072	.080	.075	.087
	N	32	32	32	32	32	32
Generation 3	Mean	.353	.353	.293	.381	.305	.309
	SD	.102	.110	.088	.075	.083	.106
	N	32	32	32	32	32	32
Generation 4	Mean	.360	.369	.272	.414	.281	.301
	SD	.122	.113	.101	.095	.089	.095
	N	32	32	32	32	32	32

Supplementary Table 7. Expt. 4 Mean proportion frequency of target-attribute pairings by Target Sex Context, Target Sex, Generation, and Attribute Type

			Female Targets			Male Targets		
			Stereotype Consistent	Stereotype Inconsistent	Neutral	Stereotype Consistent	Stereotype Inconsistent	Neutral
Single Sex Target Context	Generation 1	Mean	.320	.350	.330	.339	.328	.334
		SD	.080	.068	.073	.080	.050	.065
		N	16	16	16	16	16	16
	Generation 2	Mean	.316	.352	.331	.357	.337	.306
		SD	.098	.116	.080	.108	.075	.066
		N	16	16	16	16	16	16
	Generation 3	Mean	.305	.398	.297	.402	.307	.291
		SD	.088	.116	.094	.094	.084	.082
		N	16	16	16	16	16	16
	Generation 4	Mean	.313	.423	.264	.407	.315	.279
		SD	.094	.099	.087	.132	.102	.116
		N	16	16	16	16	16	16
Mixed Sex Target Context	Generation 1	Mean	.323	.340	.337	.376	.312	.313
		SD	.101	.108	.128	.098	.078	.072
		N	32	32	32	32	32	32
	Generation 2	Mean	.329	.343	.328	.413	.296	.292
		SD	.118	.128	.130	.125	.112	.103
		N	32	32	32	32	32	32
	Generation 3	Mean	.331	.347	.322	.430	.262	.308
		SD	.112	.152	.135	.142	.132	.127
		N	32	32	32	32	32	32
	Generation 4	Mean	.350	.331	.319	.477	.231	.293
		SD	.161	.127	.121	.136	.112	.132
		N	32	32	32	32	32	32