Research Article

Wee Kiat Lau*, Marian Sauter, Cihan Bulut, Lisa Valentina Eberhardt, Anke Huckauf

Revisiting the Watching Eyes Effect: How Emotional Expressions, Sex, and Age of Watching Eyes Influence Stereotypical Statement Endorsement

https://doi.org/10.1515/psych-2022-0135
received June 14, 2023; accepted October 17, 2023

Abstract: The watching eyes effect has been shown to influence prosocial and antisocial behaviors. However, the eye characteristics which induce this effect remain unclear. This study explored how emotional expressions (anger, fear, happy, neutral), age (old, young), and sex (male, female) of eye images affect antisocial behavior, measured by stereotype endorsement. Participants rated their endorsement of 36 stereotype statements about race, gender, and religion topics, each paired with an eye image. Our findings indicated that stereotype endorsement did not differ significantly between neutral eye images and control flower images. We then used neutral eyes as control images to evaluate stereotype endorsement when being watched by eyes with other expressions. When comparing endorsement across age and sex in each expression, the data revealed higher endorsement for angry old male eyes and for happy young eyes (males and females), and lower endorsement for fearful eyes, except for old fearful female eyes. Therefore, varying the emotional expression, sex, and age of the eye images used in the watching eyes effect paradigm influences stereotype endorsement. To further unravel the underlying mechanisms behind the watching eyes effect, we encourage future research to explore how varying characteristics of the eyes affect responses/behaviors like stereotype endorsement.

Keywords: watching eyes, emotions, sex and age

1 Introduction

The *watchful/watching eyes effect* (WEE) occurs when responses differ between conditions with a pair of eyes and conditions with a neutral image. In economic games, participants often donated more money via keypresses when a pair of stylized eyes was displayed, compared to a neutral image (Haley & Fessler, 2005; Kelsey et al., 2018; Sénémeaud et al., 2017). This led to the interpretation that a pair of eyes could influence prosocial behaviors (Haley & Fessler, 2005). Although donation behavior is a prosocial act, it can be argued that people also donate if the action follows the norm. For example, Bateson and Callow (Bateson et al., 2013) showed the WEE because there were more prosocial actions for eye images than for neutral images. They then showed an increase in prosocial actions even when the actions did not conform to the normative behaviors. Taken together, it seems advantageous to show a pair of eyes since it increases prosocial behaviors.

*** Corresponding author: Wee Kiat Lau,** Department of General Psychology, Institute of Psychology and Pedagogics, Ulm University, Albert-Einstein-Allee 47, 89081 Ulm, Germany, e-mail: wee.lau@uni-ulm.de
Marian Sauter, Cihan Bulut, Lisa Valentina Eberhardt, Anke Huckauf: Department of General Psychology, Institute of Psychology and Pedagogics, Ulm University, Albert-Einstein-Allee 47, 89081 Ulm, Germany

Several studies failed to replicate the WEE. These studies found no response differences between conditions with and without eye images (Carbon & Hesslinger, 2011; Kawamura & Kusumi, 2017; Matsugasaki et al., 2015; Sparks & Barclay, 2015). A direct imitation of the original experimental also found no evidence of the WEE (Rotella et al., 2021). Results from two meta-analyses indicated either a very weak effect or no evidence of the WEE (Nettle et al., 2013; Northover et al., 2017). Thus, what influences the response differences in the WEE appears to be inconsistent and the mechanism underlying the differences remains elusive.

It is important to scrutinize the experimental design and stimuli manipulation in the WEE experiments. Some studies have shown that the observer's environment influences how strong the WEE is. Specifically, responses differ greater for participants in crowded areas (i.e., more than three people) than for participants who are alone (Bateson et al., 2013; Kawamura & Kusumi, 2017; Nettle et al., 2013), even when the pair of eyes is always visible. Other studies show that the response differences observed in the WEE are affected by the observer's personality traits and public self-awareness (Conty et al., 2016; Pfattheicher & Keller, 2015; Rotella et al., 2021).

Dear and colleagues recently critiqued the use of prosocial tasks to investigate the WEE (2019). A person can choose not to be prosocial knowing that it will lead to negative outcomes, such as donating to criminal/questionable organizations or when the local norm frowns on prosocial behaviors. Dear and colleagues investigated the use of antisocial behaviors to investigate the WEE. Their meta-analysis found a 35% reduction in antisocial behaviors when eye images were shown. Antisocial behaviors are characterized by disruptive, harmful, or hostile actions toward others and society, with a disregard for the rights, feelings, or well-being of others. This type of behavior can manifest in diverse forms such as violence, promoting delinquency, or engaging in disruptive or criminal activities (Catalano & Hawkins, 1996). Examples of antisocial behaviors reduced by the WEE include littering (Bateson et al., 2015; Van Doesum et al., 2021), theft (Nettle et al., 2012), and lying (Oda et al., 2015).

1.1 Examining the Eye Images to Understand the Watching Eyes Effect

The eye stimulus used in the WEE experiment often depicts a pair of stylized eyes (Dear et al., 2019; Haley & Fessler, 2005; Oda et al., 2015; Rotella et al., 2021). These stylized eyes resembled the eye of Horus (Haley & Fessler, 2005). According to ancient Egyptian beliefs, the eye of Horus serves as protection to ward off danger (Darnell, 1997). However, Horus was sometimes associated with anger because of anger episodes occurring during a constant conflict with his nemesis, Seth (Balegh, 2014; Griffiths, 1958). In ancient Egypt, red was also associated with power, anger, and violence (Singer, 2010). Therefore, there is a belief that, when the eye of Horus turns red, it is a sign of anger and danger (Hussein, 2010). This also suggests that the configuration of the eye of Horus (i.e., spatial positioning of the eye region) is predisposed to expressing anger.

To date, no studies have examined if observers can perceive emotional expressions from such stylized eyes in the WEE experiment. We do not know if the response differences in the WEE are related to the perceived emotion of the eye images shown. Moreover, recent studies have adopted human eyes when investigating the WEE (Kelsey et al., 2018; Rotella et al., 2021; Shinohara & Yamamoto, 2018; Van Doesum et al., 2021), and begun questioning how features of the eye images affect the response differences in the WEE (Manesi et al., 2016; Pauwels et al., 2017; Sueur et al., 2023). We briefly describe three features that can be conveyed by the eyes: emotional expressions, age, and sex. These features, in conjunction with recent findings, may help us better understand the response differences seen in the WEE.

Human eyes convey emotional expressions. We move the eyes and the surrounding regions to show emotions, like widening the eyes to show fear or surprise (Ekman et al., 1990; Gunnery & Ruben, 2016; Hess et al., 2009; Lau, 2021; Lau & Huckauf, 2021; Lee et al., 2013; Matsumoto, 1989; Roy-Charland et al., 2015; Schützwohl & Reisenzein, 2012), or by displaying crow's feet when we make genuine smiles (Ekman et al., 1990), or by enhancing/altering eye expressions using only the eyebrows (Tipples et al., 2002). Eye sex can be estimated by looking at the eye size (Midelfart, 1996) or the skin color around the eye region (Nestor & Tarr, 2008). We may also guess how old a person is by looking at the number of wrinkles in the eye region (Ganel, 2015; Ganel & Goodale, 2021). From these, more abstract characteristics of a person, such as attractiveness,

trustworthiness, and intelligence, can be read just by looking at variations in the eyes (Kret & De Dreu, 2019; Kret et al., 2015; Lau et al., 2022; Todorov & Duchaine, 2008; Todorov et al., 2009; Van Breen et al., 2018; Willis & Todorov, 2006).

Recent studies using human eyes have unveiled some key components of why responses differ in the WEE. For instance, gaze direction refers to where the eyes are looking. Averted gazes (i.e., looking away) do not elicit response differences in the WEE compared to direct gazes (i.e., looking into the camera/observer; Manesi et al., 2016). However, several studies that used forward-gazing eyes also failed to replicate response the response differences in the WEE (Carbon & Hesslinger, 2011; Kawamura & Kusumi, 2017; Matsugasaki et al., 2015; Rotella et al., 2021; Sparks & Barclay, 2015). This means that gaze direction is not the only contributing factor. Since the eyes convey several information simultaneously, it is possible that these eye features act in combination to elicit response differences.

1.2 Known Eye Features that Influence the Watching Eyes Effect

Recent studies varied the emotional expression, age, and sex of the eye images to investigate the WEE. Emotional expressions of the eye images may shape response differences seen in the WEE. Pauwels and colleagues (2017) showed that the response differences were greater for angry eyes than for happy eyes. Specifically, participants donated and cooperated more when viewing angry eyes than happy eyes. One limitation of the study was using only young female eye images. According to the literature, how strong a facial expression is perceived depends on both the observer's and the actor's sex (Hall & Matsumoto, 2004; Johnsen et al., 1995; Matsumoto, 1989; Sacco & Hugenberg, 2009). Specific to anger and fear, anger expressions are often associated with men, and fear expressions are often associated with women (Becker et al., 2007; Hess et al., 2009). This means that specific emotional expressions and the sex of the person expressing the emotions are highly intertwined. In Pauwels and colleagues' study (Pauwels et al., 2017), female eye images were enough to demonstrate response differences in angry eyes but not in happy eyes. However, angry expressions are associated more with men, than with women (Becker et al., 2007; Hess et al., 2009). If response differences observed in the WEE are due to the perceived emotional expression from the eye images, then emotional expressions from different sexes should affect the magnitude/direction of the response differences.

Studies that use eye images from different sexes do not show response differences (Manesi et al., 2016; Nettle et al., 2013; Panagopoulos, 2014; Pauwels et al., 2017; Shinohara & Yamamoto, 2018; Sueur et al., 2023). However, the lack of response differences in the WEE occurs between neutral male and neutral female eyes. As one study has shown, eye expressions can influence the WEE (Pauwels et al., 2017). Furthermore, emotional expressions are tightly coupled to the sex of the person displaying the emotions (Hall & Matsumoto, 2004; Johnsen et al., 1995; Matsumoto, 1989; Sacco & Hugenberg, 2009). Thus, when varying eye expressions in the WEE experiment, it is also important to vary the eye sex to evaluate whether emotional expressions influence a change in responses or behaviors in the WEE paradigm. If the response differences in the WEE are due to the perceived emotional expressions, then emotional expressions from male and female eyes should be used because of the convoluted nature between emotional expressions and sex.

There are also hints that age is important to investigate in the WEE. One study showed a decrease in antisocial behaviors when viewing a child's eyes, in contrast to older eyes (Sueur et al., 2023). Specifically, participants obeyed traffic laws more after seeing a child's eyes, compared to an adult male's eyes or an adult female's eyes. This finding is not surprising since we usually consider that children are the world's future. Consequently, we do our best to ensure their survivability (Clark et al., 2020; Fulcher & Coyle, 2011). What remains unanswered are questions concerning older ages. From the literature, we commonly assume that older people hold powerful and senior positions in large organizations (Cox & Harquail, 1991; Shen & Cannella, 2002) and that they, like our parents or grandparents, have more experience and are wiser than us (Kogan & Shelton, 1962). This can also shape our respect towards the elders, even if performance does not differ across age (Jacobs et al., 1990; Yue & Ng, 1999). Since younger people often act differently and harbor different expectations about those from other age groups (Finkelstein et al., 2013; Hummert et al., 1994), it makes sense

that a person's behavior would change in the vicinity of someone else from another age group. Hence, eyes of different ages could affect how much responses change in the WEE.

1.3 Searching for Comparable Images that Contrast Eyes with Varied Expressions

The between-subject design is almost always used in past literature to evaluate the WEE (Bateson et al., 2013; Ekström, 2012; Pfattheicher & Keller, 2015). Specifically, one group of participants is shown only a pair of neutral eyes, and the other group is presented with only a neutral image, like flowers (Francey & Bergmüller, 2012; Oda & Ichihashi, 2016; Shinohara & Yamamoto, 2018). With this design, some studies did not replicate response differences (Dancer & Burn, 2019; Fenzl & Brudermann, 2021; Rotella et al., 2021). One study even found a reversed effect, where responses differed more for the flower images than for the neutral eye images (Raihani & Bshary, 2012).

Experimentally speaking, eyes and flowers have very different visual features (e.g., color, shape, size, etc.). Flowers can look more complex than eyes, depending on how complicated the petals are positioned. However, we do not detect them the same way as human eyes (McCarthy et al., 1997). Even when the flower is made to look like a pair of eyes, it neither represents human eyes, nor elicits response differences seen in the WEE (Carbon & Hesslinger, 2011). In contrast, eyes are more readily recognized since we are predisposed to recognize them almost immediately (Jessen & Grossmann, 2014; Simion et al., 2011; Simion & Giorgio, 2015). Eyes also convey social signals that are important for communication (Bavelas et al., 2002; Broz et al., 2012; Burgoon & Le Poire, 1999; Kleinke, 1986). If we varied eye features like emotional expressions, age, and sex, to investigate the WEE, then flower images would not be suitable comparisons because they differ from eyes. Instead, eye image variations should be compared to the relevant control eye image, accounting for the age and sex of the eyes.

Eyes are highly arousing stimuli (Fawcett et al., 2016; Gobel et al., 2015; Helminen et al., 2011; Jarick & Bencic, 2019). Merely becoming aware of a pair of eyes is enough to raise arousal levels (Conty et al., 2010). If arousal drives response differences in the WEE, then it makes sense to compare high-arousing eye images to low-arousing images like flowers or no image. Typically, the WEE is investigated using a between-subjects design. This design cannot account for participants' unique arousal differences since they only see an eye image instead of the whole set of images. However, arousal differences within each participant can be measured by adopting a within-subject design (Charness et al., 2012). There is only one study adopting this design to investigate the WEE (Bateson et al., 2006). Importantly, the study found an increase in cooperative behavior whenever eye images were presented compared to flower images. Since within-subject experiments also show response differences in the WEE, we can evaluate whether arousal is important by adopting a within-subject design in our study. Specifically, we would expect responses to differ between neutral eye images and the control flower images.

To better understand the role of arousal, it would be important to first show or replicate an absence of an effect when comparing neutral eyes to control flower images, following past studies (Bateson et al., 2013; Ekström, 2012; Pfattheicher & Keller, 2015). If responses differ between neutral eyes and control flower images, this suggests that arousal plays a role in the WEE. However, if there is an absence of an effect between neutral eyes and control flower images, then this suggests that the current comparison method (between eyes and flowers) cannot evaluate the effect. Then, the WEE should be further evaluated by comparing only the different eye images to determine if specific eye images (or features) induce a change in responses.

1.4 Study Goal

The following was an exploratory study. Dear and colleagues (2019) recommended investigating antisocial behaviors from their meta-analysis. We adopted their recommendation and used the extent to endorse stereotype statements as a kind of antisocial behavior (Crosby et al., 1980; Devine, 1989; Genthner & Taylor,

1973; Park et al., 2003; Zhang, 2010). Antisocial behavior was quantified by the extent participants would endorse 36 stereotypical statements about race, religion, and gender. Participants rated the extent they would endorse a given stereotype statement presented on the screen.

We first explored whether arousal between neutral eyes and flower images played a role in inducing response changes in the WEE paradigm. We hypothesized that stereotype endorsement ratings would differ between neutral eyes and control flower images. If stereotype endorsement did not differ between flower images and neutral eyes, then this would suggest that arousal between neutral eyes and flower images does not play a huge role in the WEE effect. Hence, neutral eye images would be used as control images for comparing other eye images with different characteristics.

We then explored whether the underlying mechanism behind response differences in the WEE paradigm was driven by individual, varying eye features (emotional expressions, sex, and age) or a combination of them.

We hypothesized that the stereotypical statement endorsement would be larger for emotionally expressive eyes compared to non-expressive eyes (i.e., neutral). Based on our review, emotional expressions and sex are closely related. There is a tendency to associate anger with males and fear with females (Becker et al., 2007; Hall & Matsumoto, 2004; Hess et al., 2009; Johnsen et al., 1995; Matsumoto, 1989; Sacco & Hugenberg, 2009). Thus, when varying emotional expressions, one cannot ignore how sex also influences perceived emotions. Since previous studies investigating emotional expressions in the WEE used only anger and happy expressions (Pauwels et al., 2017), we also included fear because this expression is quickly recognized (Hoffmann et al., 2010; Misailidi & Bonoti, 2008).

We hypothesized that stereotype statement endorsement would differ between old-eye images and young-eye images. This is because a child's eyes elicited the largest behavioral changes than adult eyes (Sueur et al., 2023). Furthermore, the literature also showed that people typically behave differently around those from very different age groups (Finkelstein et al., 2013; Hummert et al., 1994). We compared young adult eyes to old adult eyes since no studies have made such comparisons.

In the current study, we adopted a within-subject design to investigate whether emotional expressions, sex, and age of eye images could influence how much participants endorse 36 stereotypical statements across race, religion, and gender. From past research, forward-gaze elicits response differences in the WEE but not averted-gaze (Manesi et al., 2016). We presented colored forward-gazing human eyes. We first compared neutral eye expressions to control flower images, imitating past literature (Fenzl & Brudermann, 2021; Manesi et al., 2016), before comparing neutral eyes to eyes of other expressions. The eye expressions used were happy, angry, fear, and neutral. Eye sexes from male and female eyes were used. Eye ages from young and old adults were used. All eye images (emotional expression, sex, and age) were counterbalanced.

2 Methods

2.1 Participants

Participants were recruited in accordance with the Declaration of Helsinki. The study was approved by the ethics committee of Ulm University within the context of the DFG-funded project, grant number 425867974. Power analysis was conducted before recruitment using G*Power 3.1 (Faul et al., 2009). About 170 participants were needed for the repeated measures F-test to achieve power $(1 - \beta) = 0.99$, $\alpha = 0.05$, and small effect size $f = 0.10$. We expected a small effect size because many studies failed to replicate the watching eyes effect. Therefore, if there is indeed an effect, it would likely be a small one. From this calculation, we recruited 232 respondents between May 7 and Jun 14, 2022. All participants provided informed consent prior to the study.

Those who failed the two comprehension check questions were excluded from the analysis (shown in Design below). From the data, 23 (9.91%) failed the first comprehension check, and 44 (18.97%) failed the second comprehension check. Thus, 188 participants (mean age = 25.68, SD = 7.18) successfully completed the experiment. There were more female (N = 103, 57.79%) than male (N = 84, 44.68%) and diverse (N = 1, 0.53%)

respondents. The sampling method was of convenient means (i.e., friends, colleagues, social media). Those from Ulm University were compensated with course credits. Others volunteered for the experiment knowing that there was no monetary incentive at the end of the experiment. Thus, participants' age and gender were not controlled in the current study.

2.2 Stimuli

Eye images used in the experiment were derived from full faces of the FACES database (Ebner et al., 2010) and consent to use the images was given in writing by the database administrators. The database features Caucasian faces across various ages and sex. All images from the database are colored. The original face images have the following dimensions: 2,835 × 3,543 pixels, 300 pixels per inch, and 8 bits depth. Eight models were selected from the database (two young males, two old males, two young females, and two old females). The following IDs were selected for old (ID = 004, 146, 012, 060) and young (ID = 031, 057, 152, 182) faces. The average perceived age of these images from the FACES database (Ebner et al., 2010) was 66.87 years (SD = 3.36) for old faces and 26.44 years (SD = 2.21) for young faces. The stimulus selection criteria are further described in Appendix A1

Each model expressed angry, fear, happy, and neutral expressions. We cropped the eyes and eye region such that all eye stimuli used in the experiment had the following dimensions: 1,356 × 418 pixels, 300 pixels per in. (Figure 1a). The image manipulation was done using GIMP (version 2.10). There were four flower images which acted as control images for comparisons with neutral eyes (Figure 1b). The control images had the same dimensions as the eye stimuli. In total, we used eight models × four expressions + four flowers = 36 images for the experiment.

Stereotype statements used in the experiment were in the German language. Statements were taken from reputable sources such as news websites and the German-based market and consumer data company, Statista. Example news websites include Zeit.de, which is central leaning and does not hold strong political biases (Dallmann et al., 2015). Stereotype statements were used to poll the German public's opinions on various socio-economic and political issues between 2000 and 2020. We selected 36 statements to match the number of images used in the experiment. These statements were divided into three stereotype categories: race, gender, and religion. An example stereotypical statement concerning gender is "*Meiner Meinung nach sind Frauen hormongesteuert*" *(In my opinion, women are hormone-driven).* The full list of the statements used is found in Appendix A2

Half the stereotypical statements were phrased to convey positive meaning. This was to prevent participants from always responding using the same rating if all the statements were negatively phrased. For example, "*Meiner Meinung nach sind Frauen schlecht in Mathe*" *(In my opinion, women are bad at Math)* was changed to "*Meiner Meinung nach sind Frauen gut in Mathe*" *(In my opinion, women are good at Math).* The conversion was done by a native German speaker. We also ensured that statements across the three categories began with "Meiner Meinung/Erfahrung nach" (In my opinion/experience) whenever possible.

2.3 Design

The experiment was within-subject design involving 36 trials that corresponded to the images used. The experiment was administered online using the EFS Survey by Questback GmbH (2021). The experiment was conducted in German. All 36 images and 36 stereotype statements were fully randomized to ensure that no specific image could be attributed to a specific statement. We presented each trial on a single page. Participants always answered the question: "Wie wahrscheinlich würden Sie diese Aussage machen?" (How likely would you make this statement?) in each trial.

As the experiment was conducted in German, the question "Wie wahrscheinlich würden Sie diese Aussage machen?" was translated literally to "How likely would you make this statement?". However, terms such as "likely" or "likelihood" could carry meaning related to probability or statistics. To avoid such confusion, we used the phrase

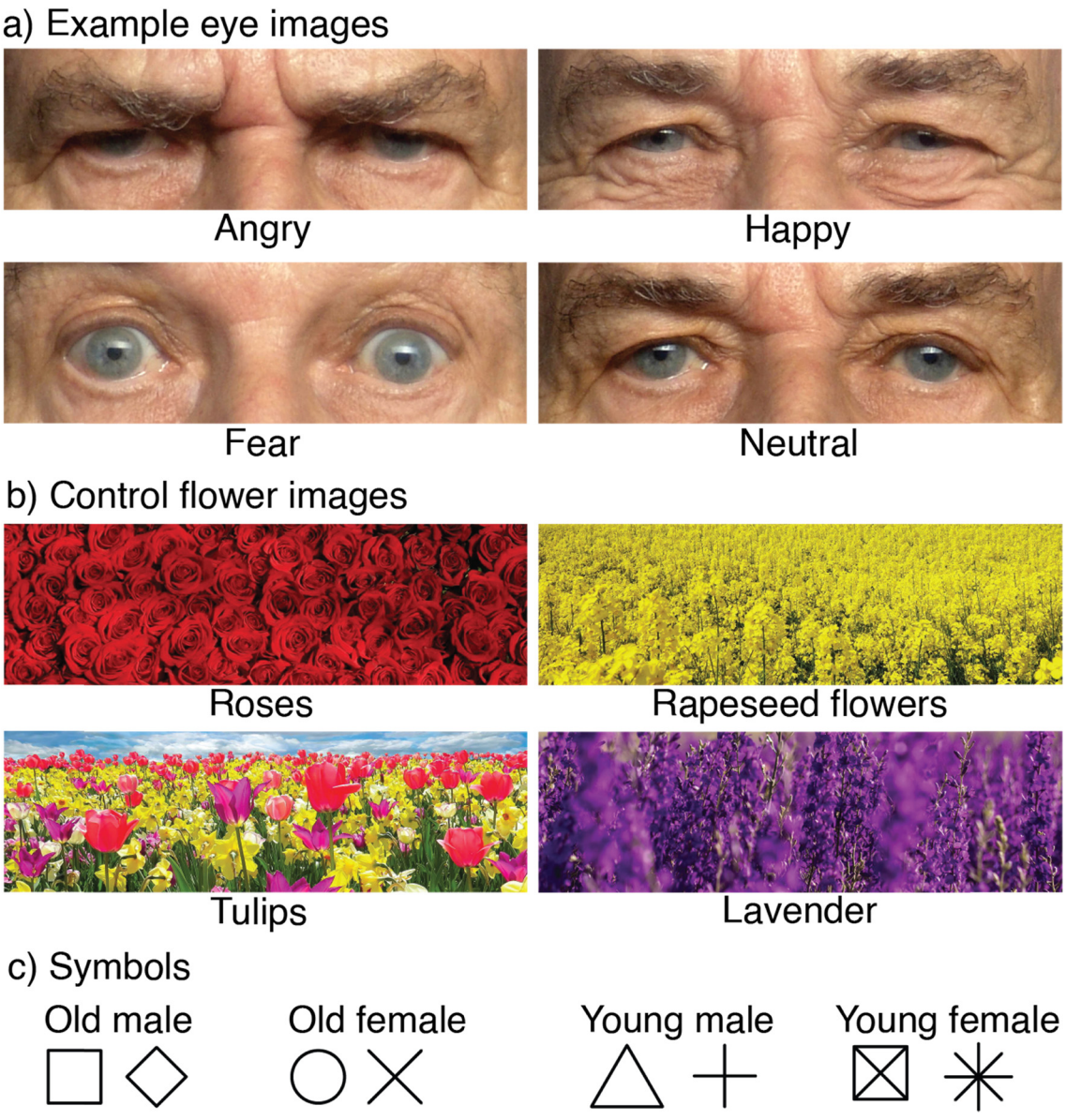

Figure 1: The figure shows (a) example eye images, (b) control flower images used in the experiment, and (c) the symbols depicting the eight eye images used in the study. The illustration for (a) depicts an old adult male expressing angry, happy, fear, and neutral expressions. Eyes were derived from full faces in the FACES database (Ebner et al., 2010). The symbols depicted in this figure apply to Figures 3 and 4 in the Results section.

"how much the participant endorses the stereotype statement" in this manuscript. The replaced phrase is to broadly refer to whether participants agree with the stereotype statements based on their own views.

A stimulus appeared only at the top of the page, center-aligned to the page. The stereotypical statement was presented below the stimulus, left-aligned to the page. The question was left-aligned to the page and was always centered between the stimulus and the statement. Figure 2a shows two example screenshots of a trial and the randomization between eye stimuli and stereotypical statements.

There were four instruction pages. On the first page, participants learned about the 9-point Likert scale used in the task. The scale indicated how much the participant would endorse the stereotype statement shown. The value 1 indicated the least and the value 9 indicated the most. A comprehension check question was implemented here. Participants had to click on "1" to indicate "the least." The second instruction page described the duration

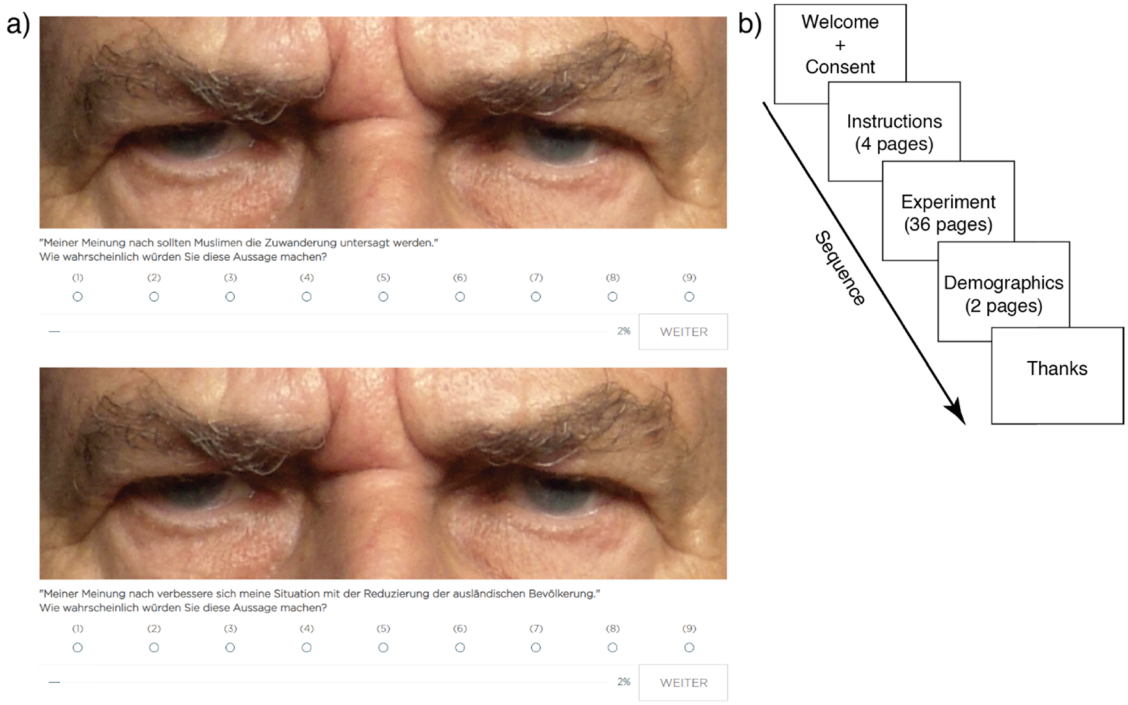

Figure 2: An example of two individual screenshots of a trial across two participants is depicted in (a). The screenshots show an eye image of an old angry male above two different stereotype statements. The experimental sequence is illustrated in (b). All images and stereotype statements were fully randomized across participants.

and number of trials. The third instruction page was an example of how each trial would look like, and where participants can indicate their responses. The fourth page was feedback on their responses.

Participants then entered brief demographic information about themselves (i.e., gender, age, ethnicity, etc.) across two pages. The second comprehension check question was implemented here. Participants must click on the corresponding number on one of the items. The experiment ended, and a thankyou screen was shown together with a brief explanation of what the experiment was about.

Respondents who completed the survey on mobile devices were encouraged to use landscape orientation. There was no backward button and returning to previous pages to change responses was not allowed. All questions must be answered during the experiment. Otherwise, the message “Bitte beantworten Sie diese Frage” (Please answer this question) appeared. Participants were neither asked whether they saw the eye images nor to report the eye emotions during the study.

There were 36 trials in the experiment, corresponding to 36 images and stereotype statements used. Each trial was presented on a separate page. In total, there were 44 pages in the experiment (Figure 2b). The entire experiment took 10–20 min to complete.

2.4 Procedure

Participants first saw the welcome screen and were encouraged to orient their personal devices to landscape mode so that the images would display properly during the experiment. Informed consent was provided by clicking on “Ich bin einverstanden” (I agree) and on the “weiter” icon at the bottom right of the page before starting the experiment.

The instructions concerning the Likert scale, the task, and how the trials would appear were shown. Participants navigated through the four instructional pages before commencing with the experimental trials. In each of the 36 trials, an image appeared on top of the page. Participants rated how likely they would say the stereotypical statement on the 9-point Likert scale. At the end of the experiment, basic demographic information (i.e., gender, age, ethnicity,

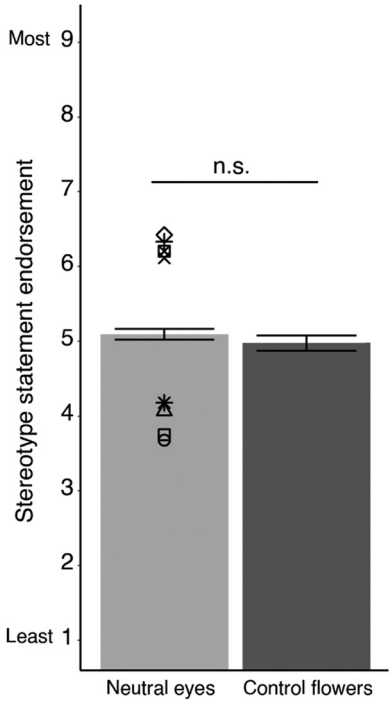

Figure 3: Bar graph comparing mean ratings between neutral eyes and control flower images Symbols indicate the averaged ratings for each eye model. Error bars illustrate standard error of means (SEM). n.s. – non-significant.

etc.) across two pages were surveyed. The experiment ended with a thankyou screen and a brief explanation of what the experiment was about.

2.5 Analysis

The analysis was conducted in SPSS Version 28.0. Those who failed the comprehension check questions were excluded from the analysis (2.1 Participants). Since we rephrased half the negative stereotypes to prevent participants from responding with the same response, we re-coded the responses of the negative stereotypes so that the arithmetic values did not neutralize themselves when we averaged participants' responses. This also eases data interpretation as higher ratings indicate higher stereotype endorsement.

We evaluated whether variations (i.e., emotional expressions, sex, and age of the watching eyes) affected stereotype statement endorsement. The dependent variable (DV) was the level of endorsement. The independent variables (IVs) were emotional expressions (angry, fearful, happy, and neutral), eye sex (male and female), and eye age (young and old).

We first compared endorsement between neutral eyes and flower images by running a paired-samples *t*-test on the DV. Since we found no differences between neutral eyes and flower images (Figure 3), we used neutral eyes as the control image and ran the remaining analyses by comparing only the eye images.

We also checked whether participants could have responded to the three stereotype categories differently by running a one-way repeated measures ANOVA with the factor: stereotype categories (race, gender, and religion). The results showed that it was unlikely that participants responded more biased toward a specific stereotype category (Appendix A3). Therefore, subsequent analyses did not involve comparing responses across different stereotype categories.

We then ran a 4 × 2 × 2 repeated measures ANOVA to test whether emotional expressions, sex, and age influenced stereotype endorsement in the WEE paradigm. The main effects for each factor, the two-way interactions (eye sex × emotional expressions, eye age × emotional expressions) and the three-way

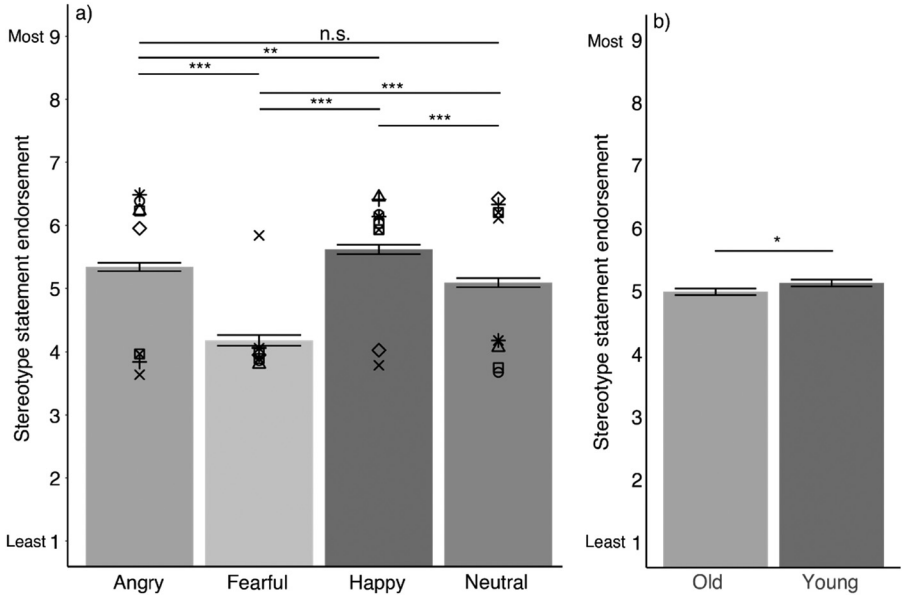

Figure 4: Bar graphs for the factors (a) eye emotional expressions, and (b) age. Symbols in (a) indicate mean responses for each eye model/identity. Error bars illustrate the standard error of means (SEM). *** $p < 0.001$, ** $p < 0.010$, * $p < 0.050$, n.s. – non-significant.

interaction (eye sex × eye age × emotional expressions) were examined. Post-hoc comparisons were conducted for significant main and interaction effects. A conservative approach to account for inflated α values was adopted via Bonferroni-correction (Abdi, 2007). For the three-way interaction, we compared emotional expressions across both ages and sexes (e.g., happy eyes: old male vs old female). We also compared each emotional expression to neutral eyes for every age and sex (e.g., old male: happy eyes vs neutral eyes). Finally, we included a comparison between young female happy eyes and young female angry eyes to confirm the findings of previous literature (Pauwels et al., 2017). Therefore, this resulted in a Bonferroni-corrected α value of 0.05 ÷ (4 expressions × 2 age × 2 sex + 3 expression pairs × 2 age × 2 sex + 1 extra pair) = 0.0017.

To streamline the readability of this manuscript, we present the results section in the following order: we first present the comparison between neutral eyes and control flowers. We then summarize the main effects of each factor (emotional expression, age, and sex) and the two-way interactions. Finally, we show results concerning the three-way interactions. We described the post-hoc comparisons for each factor with significant main effects. Since the analyses of the two-way interactions are like the three-way interactions, the post hoc results for the two-way interactions are reported in Appendix A4.

Statistical assumption checks were conducted for the repeated measures ANOVA. We checked whether the sphericity assumption was violated using Mauchly's W (Mauchly, 1940). Greenhouse–Geisser correction was applied when comparisons violated the sphericity assumption (Greenhouse & Geisser, 1959). The results of the statistical checks were reported in Appendix A5.

3 Results

3.1 No Difference Between Neutral Eyes and Flower Images

The extent to endorse stereotypical statements did not differ between neutral eyes (M = 5.09, SD = 0.98) and flower images (M = 4.97, SD = 1.38), $t(187)$ = 0.98, p = 0.330, and Cohen's d = 0.07 (Figure 3). Therefore, subsequent analyses compared only the eye images.

Table 1: Analyses of Variance for each factor and two-way interactions

Factor	Sum of squares	df	Mean square	*F*	Partial η^2
EE error	873.27	2.75	317.77	66.31***	0.26
	2462.70	513.90	4.79		
Age error	14.80	1	14.80	4.20*	0.02
	658.73	187	3.52		
Sex error	4.32	1	4.32	1.14	0.006
	708.34	187	3.79		
EE × Age error	348.04	3	116.01	33.38***	0.15
	1950.06	561	3.48		
EE × Sex error	84.17	3	28.06	8.01***	0.04
	1965.55	561	3.50		
Age × Sex error	0.83	1	0.83	0.25	0.001
	621.58	187	3.32		

Emotional expression (EE), ***$p < 0.001$, *$p < 0.05$.

3.2 Higher Stereotype Endorsement in Happy Eyes and Younger Eyes, But Lower Stereotype Endorsement in Fearful Eyes

The main effects of the factors, emotional expression, age, sex, and the two-way interactions are reported in Table 1. Details of the post-hoc comparisons for the two-way interactions are found in Appendix A4

The main effect of emotional expressions showed that they affected stereotype endorsement (Figure 4a). Post-hoc comparisons using Bonferroni-corrected $\alpha = 0.008$ revealed that the extent to endorse stereotypical

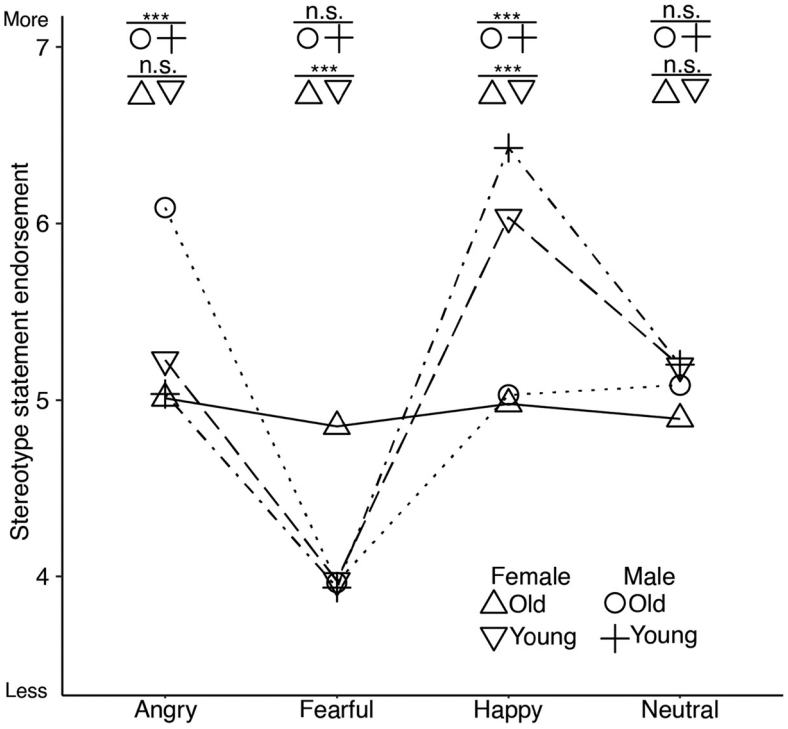

Figure 5: The three-way interaction between the factors: emotional expressions, age, and sex. Symbols indicate age and sex. *** $p < 0.001$, n.s. – non-significant.

Table 2: Post-hoc statistics of the three-way interaction emotional expression across age and sex

Emotion	Group 1		Group 2		*t*(187)	Cohen's *d*
	M	SD	*M*	SD		
	Old male		**Old female**			
Angry	6.09	2.16	5.01	1.79	5.30***	0.39
Fearful	3.97	2.09	4.85	1.95	−4.10***	−0.30
Happy	5.03	1.79	4.98	1.84	0.27	0.02
Neutral	5.09	2.02	4.89	1.74	1.02	0.07
	Young male		**Young female**			
Angry	5.03	1.74	5.23	1.90	−0.97	−0.07
Fearful	3.94	1.87	3.98	2.09	−0.23	−0.02
Happy	6.43	1.88	6.03	1.93	2.19**	0.16
Neutral	5.20	1.75	5.19	1.99	0.04	0.003
	Old male		**Young male**			
Angry	6.09	2.16	5.03	1.74	5.03***	0.37
Fearful	3.97	2.09	3.94	1.87	−7.64***	−0.56
Happy	5.03	1.79	6.43	1.88	0.16	0.01
Neutral	5.09	2.02	5.20	1.75	−0.6	−0.04
	Old female		**Young female**			
Angry	5.01	1.79	5.23	1.90	−1.17	−0.09
Fearful	4.85	1.95	3.98	2.09	4.42***	0.32
Happy	4.98	1.84	6.03	1.93	−5.55***	−0.40
Neutral	4.89	1.74	5.19	1.99	−1.54	−0.11

***$p < 0.001$, **$p < 0.01$.

statements was higher for angry eyes ($M = 5.34$, SD = 0.91), than fearful eyes ($M = 4.18$, SD = 1.14), $t(187) = 10.46$, $p < 0.001$, and Cohen's $d = 0.76$. The extent was also less in angry eyes than in happy eyes ($M = 5.62$, SD = 1.03), $t(187) = -2.87$, $p = 0.005$, and Cohen's $d = -0.21$. There was no difference between angry eyes and neutral eyes ($M = 5.09$, SD = 1.28), $t(187) = 2.66$, $p > 0.008$, and Cohen's $d = 0.19$.

For fearful eyes, participants showed the least stereotypical statement endorsement. That is, ratings were significantly lower than for happy watching eyes, $t(187) = -11.32$, $p < 0.001$, Cohen's $d = -0.83$, and neutral eyes, $t(187) = -8.35$, $p < 0.001$, and Cohen's $d = -0.61$. For happy eyes, participants were more inclined to endorse stereotypical statements than for neutral eyes, $t(187) = 4.85$, $p < 0.001$, and Cohen's $d = 0.35$.

There was a significant main effect of eye age on stereotypical statement endorsement, $F(1, 187) = 4.20$, $p = 0.040$, and $\eta_p^2 = 0.02$ (Figure 4b). The extent was lower for old eyes ($M = 4.99$, SD = 0.67) than for young eyes ($M = 5.13$, SD = 0.64). There was no significant main effect of eye sex on making stereotype statements, $F(1, 187) = 1.14$, $p = 0.290$, and $\eta_p^2 = 0.006$.

In summary, we found the effects of emotional expressions and age of the watching eyes on the extent to endorse stereotypical statements. Happy eyes, along with young eyes, influenced the most stereotypical statement endorsement. Fearful eyes influenced the least stereotype endorsement.

3.3 Higher Stereotype Endorsement for Old Angry Male Eyes, Happy Eyes, and Old Fearful Female Eyes, Compared to Each Respective Expression

The goal of the three-way interactions was to further explore specific eye features used in the study that contributed to the main effects and two-way interactions. From the analysis, we found a significant three-way interaction between emotional expressions, age, and sex, $F(3, 561) = 11.23$, $p < 0.001$, and $\eta_p^2 = 0.06$ (Figure 5). Post-hoc comparisons using Bonferroni-corrected $\alpha = 0.0018$ are reported in Tables 2 and 3.

Table 3: Post-hoc statistics of each eye expression against neutral eyes per age and sex

Eye image	Emotion		Neutral		*t*(187)	Cohen's *d*
	M	SD	*M*	SD		
Old male	Angry					
	6.09	2.16	5.09	2.02	4.92***	0.36
	Fearful					
	3.97	2.09			−5.33***	−0.39
	Happy					
	5.03	1.79			−0.29	−0.02
Old female	Angry					
	5.01	1.79	4.89	1.74	0.66	0.05
	Fearful					
	4.85	1.95			−0.22	−0.02
	Happy					
	4.98	1.84			0.45	0.03
Young male	Angry					
	5.03	1.74	5.20	1.75	−0.95	−0.07
	Fearful					
	3.94	1.87			−6.62***	−0.48
	Happy					
	6.43	1.88			6.48***	0.47
Young female	Angry					
	5.23	1.90	5.19	1.99	0.17	0.01
	Fearful					
	3.98	2.09			−6.00***	−0.44
	Happy					
	6.03	1.93			4.09***	0.30

***$p < 0.001$.

Table 2 depicts the post-hoc statistics of each emotional expression between the factors of age and sex. Table 3 shows the post-hoc statistics of each emotional expression compared to neutral eyes for each age and sex.

We then evaluated whether our data supported an existing literature showing higher response tendencies for angry young female eyes than for happy young female eyes (Pauwels et al., 2017). Our data showed similar response patterns. In our study, how much the participant endorses the stereotypical statements was higher in angry female eyes than in happy female eyes, $t(187) = -4.03$, $p = < 0.001$, and Cohen's $d = -0.29$.

Summarizing the results from the three-way interaction, the data revealed that the extent to endorse stereotypical statements was dependent on the kinds of emotional expressions, the age, and the sex of the eyes. Old angry eyes elicited a larger endorsement for stereotypical statements in male eyes. Fearful eyes elicited a stronger endorsement for stereotypical statements in old female eyes. Across both sexes, happy eyes elicited a larger extent to endorse stereotypical statements for young compared to old eyes.

4 Discussion

Only a handful of studies have investigated whether the kinds of watching eyes influence response differences in the WEE (Manesi et al., 2016; Pauwels et al., 2017; Sueur et al., 2023). In the present study, we explored whether varying emotional expression, age, and sex of the eye images affected stereotypical statement endorsement. Our hypothesis that arousal was not a mechanism behind the WEE was supported, replicating past studies (Dancer & Burn, 2019; Fenzl & Brudermann, 2021; Rotella et al., 2021). Our hypothesis about differences in endorsements between emotionally expressive eyes and neutral eyes was also

supported, indicating that future studies could compare neutral eyes with other eye expressions in the WEE paradigm instead (Carbon & Hesslinger, 2011; Kawamura & Kusumi, 2017; Matsugasaki et al., 2015; Northover et al., 2017; Rotella et al., 2021; Sparks & Barclay, 2015). The hypothesis that stereotype endorsements would differ between old eyes and young eyes was supported, replicating past research (Sueur et al., 2023). What was interesting from our exploratory analysis and was the differences in endorsements between different eye expressions, ages, and sex. Specifically, we found higher stereotypical statement endorsement when showing eyes from angry old men and young happy eyes across both sexes. In contrast, endorsement was lower in fearful-watching eyes. Therefore, our findings suggest that angry and happy-watching eyes may be used to promote stereotypical statement endorsement, while fearful-watching eyes may help reduce endorsement.

4.1 The Role of Arousal in the Watching Eyes Effect Paradigm

Our hypothesis that the WEE was not associated with arousal was supported. We replicated recent studies showing no rating differences between neutral eyes and control flowers (Carbon & Hesslinger, 2011; Fenzl & Brudermann, 2021; Raihani & Bshary, 2012) even when adopting a within-subject design. This provided evidence that arousal was unlikely an important factor behind the WEE, supporting past studies. For example, Carbon and Hesslinger (2011) postulated that if the WEE was associated with arousal, then the effect should be observable when showing other arousing non-eye images, but they found no such evidence. Fenzl and Brudermann (2021) also found no evidence of WEE because participants' self-reported arousal ratings did not differ when viewing the eye image or the control flowers image.

Arousal from eye images is not as arousing as looking at real pairs of eyes. Although the eyes are highly arousing stimuli (Fawcett et al., 2016; Gobel et al., 2015; Helminen et al., 2011; Jarick & Bencic, 2019), we are bad at judging how aroused we feel when looking at static eye images partially occluded faces (Adolphs et al., 1999; Lau, 2021; Lau et al., 2022; Sato & Yoshikawa, 2007; Wang et al., 2018). We also do not experience the same physiological arousal when making eye contact with a person on a video recording, compared to making eye contact with a live person (Lyyra et al., 2018). This means that a better way of inducing arousal, such as making eye contact, is required to determine whether arousal plays a role in the WEE.

What remains unclear is also the psychophysical data about the observers' arousal levels. Studies show a drastic increase in arousal when monitoring skin conductance and heart rate as observers make eye contact with another person (Helminen et al., 2011; Jarick & Bencic, 2019; Nichols & Champness, 1971). However, the arousal levels in healthy participants are lower when viewing images (Wang et al., 2018). No studies have measured participants' internal arousal levels during a WEE experiment. Thus, there is still a need to rule out whether participants' own arousal in the experiment influences their responses.

4.2 Happy Eyes Trigger a Sense of Acceptance

Our hypothesis that stereotype endorsement would differ between emotionally expressive eye images and neutral eye images was also supported. In our study, participants gave high endorsements when viewing happy young eyes.

Positive valence, such as feeling happy or seeing happy eyes, may trigger a sense of acceptance from social norms. Affective valence, that is, how positive or negative we feel, influences the way we interact with objects or people. Everyday objects elicit some form of valence (Lebrecht et al., 2012). We are more likely to approach positive valence stimuli and avoid negative valence stimuli (Alexopoulos & Ric, 2007). Nettle and colleagues (Nettle et al., 2013) suggested that an alternative explanation for the WEE could be that people are just doing things that conform to the social norms. If watching eyes make people act more closely to the social norms, then the extent to endorse stereotypes should increase if the norm reinforces such stereotypes. In our study, we

observed high endorsement scores for happy young eyes. This could be interpreted as happy eyes inducing a sense of approval, and whichever response the participants made were seen as acceptable by the social norm.

The absence of an approval/acceptance effect for old happy eyes could be attributed to own-age face biases. In face perception, own-age faces are remembered and recognized more accurately (Anastasi & Rhodes, 2005; Harrison & Hole, 2009; Rodin, 1987). This means that younger participants are more likely to recognize expressions in younger faces. Since our participants were mostly young adults, they would be better at recognizing young happy eyes than old happy eyes. Thus, participants would not have felt a sense of approval for endorsing stereotype statements when viewing old happy eyes if participants did not strongly perceive positive valence from those eyes.

4.3 Fearful Eyes Promote Avoidance from the Content

We found that stereotypical statement endorsement was lower across almost all fearful eyes, despite randomizing the presentation of the statements (Figure 4). It was possible that the fearful watching eyes signaled warning/danger against endorsing stereotypes (Jack et al., 2014; Mathews et al., 2003; Morris et al., 2002). Fear is a powerful emotion because it acts as a defense mechanism for protection (LeDoux, 2012). We perceive whether a person is feeling afraid because fear is conveyed through facial expressions and body language (De Gelder, 2006). Fear can already be recognized from raised eyebrows and widened eyes (Dadds et al., 2006; Kohler et al., 2004; Matsumoto, 1989; Vernon & Berenbaum, 2002).

It is common for observers to react to another person exhibiting fear (Marsh & Ambady, 2007). Such reactions for the observers, called social fear, can be as strong as the person exhibiting the fear (Olsson et al., 2007). In fact, observers avoid fear-inducing stimuli when the stimuli cause fear in another person (Debiec & Olsson, 2017). Since messages from a fearful person appear more reliable and believable than those from a neutral person (Reed & DeScioli, 2017), this indicates that the contents experienced by a fearful person serve as an indicator of avoidance.

Fear as an indicator of avoidance could explain why participants gave the least endorsement when viewing fearful eyes. Specifically, participants did not endorse statements presented together with fearful watching eyes due to fear avoidance (Debiec & Olsson, 2017; Haaker et al., 2017; Olsson et al., 2007). Participants might have distanced themselves from the stereotypical statements when viewing fearful eyes. Therefore, we speculate that fearful-watching eyes may be used in future studies to reduce or deter stereotype endorsements.

4.4 Final Note on Ethical Concerns

It is important to point out the potential ethical concerns surrounding the study. We showed that stereotype endorsement was higher when displaying angry old male eyes and young happy eyes. We also prompted future research to use stimuli with similar features to investigate stereotype endorsement in the WEE paradigm. However, this can be ethically challenging when an experiment promotes antisocial/maladaptive behaviors like reinforcing stereotypes. Although there is a need to understand what evokes the response differences in the WEE, we caution researchers to carefully measure the risks and benefits of promoting certain behaviors when examining the WEE.

4.5 Future Directions

The current study may benefit from further improvements. It is unknown whether participants knew about the nature of the study. We do not know where participants align themselves when endorsing stereotypical

statements. Future research could investigate if participants' prior expectations of the study goal could bias the results or formulate a way to measure the ground truth to evaluate stereotype endorsement. It is also unclear whether emotionally expressive eyes can influence prosocial behaviors. Specifically, we do not know if expressive eyes affect donation behaviors (Haley & Fessler, 2005). Future studies could vary the eye expression, along with the eye's age and sex, to better understand how variations of the eye stimuli affect donation amount, and/or donation frequency. In addition, ethnicity could play a role in stereotype endorsement and facial expression perception. Future studies should also investigate how ethnicity, prior stereotype biases, and emotion perception interact with eye images of varying expressions.

There are also challenges in interpreting some results. For instance, why did angry old male eyes lead to higher endorsement scores? One reason could be the daily experiences of growing up and living in a patriarchal-dominating society, in which the sight of an angry old man could conjure uneasiness and evoke behaviors to appease the angry old man. However, one might look at the results differently and ask: why were there no endorsement changes in old female eyes? Could old females be perceived as less strict? Since these interpretations are all very speculative, it would be beneficial for future studies to replicate the current findings before delving deeper into the relationship between eye age and stereotype endorsements, and further investigating how eye age influences responses in the WEE experiments.

Many factors beyond varying the eye images can also influence how responses differ in the WEE paradigm. For instance, self-referential processing when seeing a pair of eyes causes us to be more self-aware and cautious of our actions (Hietanen & Hietanen, 2017). Consequently, this motivates our need to seek positive approval of our actions (Conty et al., 2016) such as by acting more charitable (Cañigueral & Hamilton, 2019). Previous studies have also shown that a pair of eyes influences responses because of the observer's sense of public self-awareness (Pfattheicher & Keller, 2015) and the number of people in the room where the experiment took place (Oda & Ichihashi, 2016). We also did not survey our participants whether they felt watched or observed by the eye images. Future studies may consider including some or all these measures to better understand how responses change in the WEE paradigm.

Going beyond self-report, future WEE experiments should also include psychophysiological data to gain insight into the observer's potential changes of short-term phasic arousal and where the observer is looking. Examples of such data include the participant's eye movements, pupillary changes, heart rate, and skin-conductance level. Such data is needed because looking at the eyes is arousing, more so when eye contact with another person is involved (Myllyneva & Hietanen, 2015; Nichols & Champness, 1971). However, no studies have measured the observer's physiological arousal levels, aside from the self-reported arousal levels (i.e., arousal ratings). Furthermore, it is unknown how often observers look at the eyes, or do they simply ignore them. These additional parameters allow us to dissociate the role of the observer's arousal in the WEE. Replication of the data in a lab enabling collection of more behavioral data such as time spent on each page, gaze spent on the eyes and the statements, would be helpful to get a richer picture of the processes underlying the response differences.

5 Conclusion

Many studies failed to replicate the watching eyes effect when comparing neutral eyes to flower images. In this study, we explored whether the kinds of eyes shown played a role in endorsing stereotypical statements. We replicated the literature, showing no response differences between neutral eyes and flower images. Importantly, we found that emotional eyes can, depending on their sex and gender, affect how much we endorsed stereotypical statements in both directions: Old angry male eyes and young happy eyes promoted the extent to endorse stereotypical statements. In comparison, fearful eyes reduced stereotype endorsement. Our findings indicate that future studies may consider using old angry male eyes or young happy eyes to promote stereotypical endorsement and fearful eyes to deter stereotypical endorsement. Therefore, the novelty of our study shows that variations in eye images can influence stereotype endorsement. However, future research would need to further disentangle how (and possibly

why) variations in the eye images used in a watching eyes effect paradigm influence stereotypical endorsement.

Acknowledgements: The authors would like to thank all participants for giving their time and energy in making this study possible.

Funding information: This research was funded by Deutsche Forschungsgemeinschaft (DFG, German Research Foundation), grant number "425867974" and is part of the "Priority Program SPP2199 Scalable Interaction Paradigms for Pervasive Computing Environments."

Author contributions: WL: Conceptualization, Methodology, Analysis, Data Interpretation, Initial manuscript preparation, editing. MS: Data interpretation, reviewing, editing. CB: Methodology, Data collection, reviewing. LE: Data interpretation, reviewing, editing. AH: Data interpretation, reviewing, editing, resources, funding acquisition.

Conflict of interest: The authors declare no conflict of interest.

Institutional review board statement: The study was conducted in accordance with the Declaration of Helsinki, and approved by the ethics committee of Ulm University, under the DFG-funded project 425867974.

Informed consent statement: Informed consent was obtained from all subjects involved in the study.

Data availability statement: The pseudo-anonymized experimental data, including the pre-processing scripts and analysis scripts are found at http://dx.doi.org/10.18725/OPARU-46138.

References

Abdi, H. (2007). Bonferroni test. In N. J. Salkind (Ed.), *Encyclopedia of measurement and statistics* (pp. 103–107). Thousand Oaks, CA: Sage Publications, Inc. doi: 10.4135/9781412952644.

Adolphs, R., Russell, J. A., & Tranel, D. (1999). A role for the human amygdala in recognizing emotional arousal from unpleasant stimuli. *Psychological Science*, *10*(2), 167–171.

Alexopoulos, T., & Ric, F. (2007). The evaluation-behavior link: Direct and beyond valence. *Journal of Experimental Social Psychology*, *43*(6), 1010–1016. doi: 10.1016/j.jesp.2006.10.017.

Anastasi, J. S., & Rhodes, M. G. (2005). An own-age bias in face recognition for children and older adults. *Psychonomic Bulletin & Review*, *12*(6), 1043–1047. doi: 10.3758/BF03206441.

Balegh, R. (2014). Emotions and Deities in Ancient Egypt. *Studies in the Antiquities of the Arab World*. *17*(17), 49–83. doi: 10.21608/cguaa.2014.44335.

Bateson, M., Callow, L., Holmes, J. R., Redmond Roche, M. L., & Nettle, D. (2013). Do Images of 'Watching Eyes' Induce Behaviour That Is More Pro-Social or More Normative? A Field Experiment on Littering. *PLOS ONE*, *8*(12), e82055. doi: 10.1371/journal.pone.0082055.

Bateson, M., Nettle, D., & Roberts, G. (2006). Cues of being watched enhance cooperation in a real-world setting. *Biology Letters*, *2*(3), 412–414. doi: 10.1098/rsbl.2006.0509.

Bateson, M., Robinson, R., Abayomi-Cole, T., Greenlees, J., O'Connor, A., & Nettle, D. (2015). Watching eyes on potential litter can reduce littering: Evidence from two field experiments. *PeerJ*, *3*, e1443. doi: 10.7717/peerj.1443.

Bavelas, J. B., Coates, L., & Johnson, T. (2002). Listener Responses as a Collaborative Process: The Role of Gaze. *Journal of Communication*, *52*(3), 566–580. doi: 10.1111/j.1460-2466.2002.tb02562.x.

Becker, D. V., Kenrick, D. T., Neuberg, S. L., Blackwell, K. C., & Smith, D. M. (2007). The confounded nature of angry men and happy women. *Journal of Personality and Social Psychology*, *92*, 179–190. doi: 10.1037/0022-3514.92.2.179.

Broz, F., Lehmann, H., Nehaniv, C. L., & Dautenhahn, K. (2012, 9–13 Sept). *Mutual gaze, personality, and familiarity: Dual eye-tracking during conversation*. Paper presented at the 2012 IEEE RO-MAN: The 21st IEEE International Symposium on Robot and Human Interactive Communication.

Burgoon, J. K., & Le Poire, B. A. (1999). Nonverbal cues and interpersonal judgments: Participant and observer perceptions of intimacy, dominance, composure, and formality. *Communication Monographs*, *66*(2), 105–124. doi: 10.1080/03637759909376467.

Cañigueral, R., & Hamilton, A. F. d. C. (2019). Effects of being watched on self-referential processing, self-awareness and prosocial behaviour. *Consciousness and Cognition*, *76*, 102830.

Carbon, C.-C., & Hesslinger, V. M. (2011). Bateson et al.'s (2006) cues-of-being-watched paradigm revisited. *Swiss Journal of Psychology/ Schweizerische Zeitschrift für Psychologie/Revue Suisse de Psychologie*, *70*(4), 203–210. doi: 10.1024/1421-0185/a000058.

Catalano, R. F., & Hawkins, J. D. (1996). The social development model: A theory of antisocial behavior. In *Delinquency and crime: Current theories* (pp. 149–197). New York, NY, US: Cambridge University Press.

Charness, G., Gneezy, U., & Kuhn, M. A. (2012). Experimental methods: Between-subject and within-subject design. *Journal of Economic Behavior & Organization*, *81*(1), 1–8. doi: 10.1016/j.jebo.2011.08.009.

Clark, H., Coll-Seck, A. M., Banerjee, A., Peterson, S., Dalglish, S. L., Ameratunga, S., Balabanova, D., Bhan, M. K., Bhutta, Z. A., Borrazzo, J., & Claeson, M. (2020). A future for the world's children? A WHO–UNICEF–Lancet Commission. *The Lancet*, *395*(10224), 605–658.

Conty, L., George, N., & Hietanen, J. K. (2016). Watching Eyes effects: When others meet the self. *Consciousness and Cognition*, *45*, 184–197. doi: 10.1016/j.concog.2016.08.016.

Conty, L., Russo, M., Loehr, V., Hugueville, L., Barbu, S., Huguet, P., Tijus, C., & George, N. (2010). The mere perception of eye contact increases arousal during a word-spelling task. *Social Neuroscience*, *5*(2), 171–186. doi: 10.1080/17470910903227507.

Cox, T. H., & Harquail, C. V. (1991). Career paths and career success in the early career stages of male and female MBAs. *Journal of Vocational Behavior*, *39*(1), 54–75. doi: 10.1016/0001-8791(91)90004-6.

Crosby, F., Bromley, S., & Saxe, L. (1980). Recent unobtrusive studies of Black and White discrimination and prejudice: A literature review. *Psychological Bulletin*, *87*(3), 546–563. doi: 10.1037/0033-2909.87.3.546.

Dadds, M. R., Perry, Y., Hawes, D. J., Merz, S., Riddell, A. C., Haines, D. J., Solak, E., & Abeygunawardane, A. I. (2006). Attention to the eyes and fear-recognition deficits in child psychopathy. *The British Journal of Psychiatry*, *189*(3), 280–281.

Dallmann, A., Lemmerich, F., Zoller, D., & Hotho, A. (2015). *Media Bias in German Online Newspapers*. Paper presented at the Proceedings of the 26th ACM Conference on Hypertext & Social Media, Guzelyurt, Northern Cyprus. 10.1145/2700171.2791057.

Dancer, A. M. M., & Burn, C. C. (2019). Visitor effects on zoo-housed Sulawesi crested macaque (Macaca nigra) behaviour: Can signs with 'watching eyes' requesting quietness help? *Applied Animal Behaviour Science*, *211*, 88–94. doi: 10.1016/j.applanim.2018.12.005.

Darnell, J. C. (1997). The Apotropaic Goddess in the Eye. *Studien zur Altägyptischen Kultur*, *24*, 35–48.

De Gelder, B. (2006). Towards the neurobiology of emotional body language. *Nature Reviews Neuroscience*, *7*(3), 242–249.

Dear, K., Dutton, K., & Fox, E. (2019). Do 'watching eyes' influence antisocial behavior? A systematic review & meta-analysis. *Evolution and Human Behavior*, *40*(3), 269–280. doi: 10.1016/j.evolhumbehav.2019.01.006.

Debiec, J., & Olsson, A. (2017). Social fear learning: From animal models to human function. *Trends in Cognitive Sciences*, *21*(7), 546–555. doi: 10.1016/j.tics.2017.04.010.

Devine, P. G. (1989). Stereotypes and prejudice: Their automatic and controlled components. *Journal of Personality and Social Psychology*, *56*(1), 5–18. doi: 10.1037/0022-3514.56.1.5.

Ebner, N. C., Riediger, M., & Lindenberger, U. (2010). FACES – A database of facial expressions in young, middle-aged, and older women and men: Development and validation. *Behavior Research Methods*, *42*(1), 351–362. doi: 10.3758/BRM.42.1.351.

Ekman, P., Davidson, R. J., & Friesen, W. V. (1990). The Duchenne smile: Emotional expression and brain physiology: II. *Journal of Personality and Social Psychology*, *58*(2), 342. doi: 10.1037/0022-3514.58.2.342.

Ekström, M. (2012). Do watching eyes affect charitable giving? Evidence from a field experiment. *Experimental Economics*, *15*(3), 530–546. doi: 10.1007/s10683-011-9312-6.

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G* Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, *41*(4), 1149–1160. doi: 10.3758/BRM.41.4.1149.

Fawcett, C., Wesevich, V., & Gredebäck, G. (2016). Pupillary contagion in infancy: Evidence for spontaneous transfer of arousal. *Psychological Science*, *27*(7), 997–1003. doi: 10.1177/0956797616643924.

Fenzl, T., & Brudermann, T. (2021). Eye cues increase cooperation in the dictator game under physical attendance of a recipient, but not for all. *Journal of Behavioral and Experimental Economics*, *94*, 101748. doi: 10.1016/j.socec.2021.101748.

Finkelstein, L. M., Ryan, K. M., & King, E. B. (2013). What do the young (old) people think of me? Content and accuracy of age-based metastereotypes. *European Journal of Work and Organizational Psychology*, *22*(6), 633–657. doi: 10.1080/1359432X.2012.673279.

Francey, D., & Bergmüller, R. (2012). Images of eyes enhance investments in a real-life public good. *PLOS ONE*, *7*(5), e37397. doi: 10.1371/journal.pone.0037397.

Fulcher, M., & Coyle, E. F. (2011). Breadwinner and caregiver: A cross-sectional analysis of children's and emerging adults' visions of their future family roles. *British journal of developmental psychology*, *29*(2), 330–346. doi: 10.1111/j.2044-835X.2011.02026.x.

Ganel, T. (2015). Smiling makes you look older. *Psychonomic Bulletin & Review*, *22*(6), 1671–1677. doi: 10.3758/s13423-015-0822-7.

Ganel, T., & Goodale, M. A. (2021). The effect of smiling on the perceived age of male and female faces across the lifespan. *Scientific Reports*, *11*(1), 1–10. doi: 10.1038/s41598-021-02380-2.

Genthner, R. W., & Taylor, S. P. (1973). Physical aggression as a function of racial prejudice and the race of the target. *Journal of Personality and Social Psychology*, *27*(2), 207–210. doi: 10.1037/h0034776.

GmbH, Q. (2021). EFS Survey. In *Version Fall 2020*. Cologne: Questback GmbH.

Gobel, M. S., Kim, H. S., & Richardson, D. C. (2015). The dual function of social gaze. *Cognition*, *136*, 359–364. doi: 10.1016/j.cognition.2014.11.040.

Greenhouse, S. W., & Geisser, S. (1959). On methods in the analysis of profile data. *Psychometrika*, *24*(2), 95–112. doi: 10.1007/BF02289823.

Griffiths, J. G. (1958). Remarks on the Mythology of the Eyes of Horus. *Chronique d'Egypte*, *33*(66), 182–193. doi: 10.1484/J.CDE.2.309443.

Gunnery, S. D., & Ruben, M. A. (2016). Perceptions of Duchenne and non-Duchenne smiles: A meta-analysis. *Cognition and Emotion*, *30*(3), 501–515. doi: 10.1080/02699931.2015.1018817.

Haaker, J., Golkar, A., Selbing, I., & Olsson, A. (2017). Assessment of social transmission of threats in humans using observational fear conditioning. *Nature Protocols*, *12*(7), 1378–1386. doi: 10.1038/nprot.2017.027.

Haley, K. J., & Fessler, D. M. T. (2005). Nobody's watching?: Subtle cues affect generosity in an anonymous economic game. *Evolution and Human Behavior*, *26*(3), 245–256. doi: 10.1016/j.evolhumbehav.2005.01.002.

Hall, J. A., & Matsumoto, D. (2004). Gender differences in judgments of multiple emotions from facial expressions. *Emotion*, *4*(2), 201. doi: 10.1037/1528-3542.4.2.201.

Harrison, V., & Hole, G. J. (2009). Evidence for a contact-based explanation of the own-age bias in face recognition. *Psychonomic Bulletin & Review*, *16*(2), 264–269. doi: 10.3758/PBR.16.2.264.

Helminen, T. M., Kaasinen, S. M., & Hietanen, J. K. (2011). Eye contact and arousal: The effects of stimulus duration. *Biological Psychology*, *88*(1), 124–130.

Hess, U., Adams, R. B. Jr., Grammer, K., & Kleck, R. E. (2009). Face gender and emotion expression: Are angry women more like men? *Journal of Vision*, *9*(12), 19–19. doi: 10.1167/9.12.19.

Hietanen, J. O., & Hietanen, J. K. (2017). Genuine eye contact elicits self-referential processing. *Consciousness and Cognition*, *51*, 100–115. doi: 10.1016/j.concog.2017.01.019.

Hoffmann, H., Kessler, H., Eppel, T., Rukavina, S., & Traue, H. C. (2010). Expression intensity, gender and facial emotion recognition: Women recognize only subtle facial emotions better than men. *Acta psychologica*, *135*(3), 278–283. doi: 10.1016/j.actpsy.2010.07.012.

Hummert, M. L., Garstka, T. A., Shaner, J. L., & Strahm, S. (1994). Stereotypes of the elderly held by young, middle-aged, and elderly adults. *Journal of Gerontology*, *49*(5), P240–P249. doi: 10.1093/geronj/49.5.P240.

Hussein, A. M. (2010). Beware of the Red-Eyed Horus: The Significance of Carnelian in Egyptian Royal Jewelry. In Z. Hawass, P. Der Manuelian, & R. B. Hussein (Eds.), *Perspectives on Ancient Egypt–Studies in honor of Edward Brovarski.* (pp. 185–190). Egypt: Conseil suprême des Antiquités égyptiennes.

Jack, R. E., Garrod, O. G., & Schyns, P. G. (2014). Dynamic facial expressions of emotion transmit an evolving hierarchy of signals over time. *Current Biology*, *24*(2), 187–192. doi: 10.1016/j.cub.2013.11.064.

Jacobs, R., Hofmann, D. A., & Kriska, S. D. (1990). Performance and seniority. *Human Performance*, *3*(2), 107–121. doi: 10.1207/s15327043hup0302_3.

Jarick, M., & Bencic, R. (2019). Eye contact is a two-way street: Arousal is elicited by the sending and receiving of eye gaze information. *Frontiers in Psychology*, *10*, 1262.

Jessen, S., & Grossmann, T. (2014). Unconscious discrimination of social cues from eye whites in infants. *Proceedings of the National Academy of Sciences*, 111(45), 16208–16213.

Johnsen, B. H., Thayer, J. F., & Hugdahl, K. (1995). Affective judgment of the Ekman faces: A dimensional approach. *Journal of Psychophysiology*, *9*(3), 193–202

Kawamura, Y., & Kusumi, T. (2017). The norm-dependent effect of watching eyes on donation. *Evolution and Human Behavior*, *38*(5), 659–666. doi: 10.1016/j.evolhumbehav.2017.05.003.

Kelsey, C., Vaish, A., & Grossmann, T. (2018). Eyes, more than other facial features, enhance real-world donation behavior. *Human Nature*, *29*(4), 390–401. doi: 10.1007/s12110-018-9327-1.

Kleinke, C. L. (1986). Gaze and eye contact: A research review. *Psychological Bulletin*, *100*(1), 78–100. doi: 10.1037/0033-2909.100.1.78.

Kogan, N., & Shelton, F. C. (1962). Images of "old people" and "people in general" in an older sample. *The Journal of Genetic Psychology*, *100*(1), 3. doi: 10.1080/00221325.1962.10533569.

Kohler, C. G., Turner, T., Stolar, N. M., Bilker, W. B., Brensinger, C. M., Gur, R. E., & Gur, R. C. (2004). Differences in facial expressions of four universal emotions. *Psychiatry Research*, *128*(3), 235–244. doi: 10.1016/j.psychres.2004.07.003.

Kret, M., & De Dreu, C. K. (2019). The power of pupil size in establishing trust and reciprocity. *Journal of Experimental Psychology: General*, *148*(8), 1299. doi: 10.1037/xge0000508.

Kret, M., Fischer, A., & De Dreu, C. K. (2015). Pupil mimicry correlates with trust in in-group partners with dilating pupils. *Psychological Science*, *26*(9), 1401–1410. doi: 10.1177/0956797615588306.

Lau, W. K. (2021). Face masks bolsters the characteristics from looking at a face even when facial expressions are impaired. *Frontiers in Psychology*, *12*(5912), 1–14. doi: 10.3389/fpsyg.2021.704916.

Lau, W. K., Chalupny, J., Grote, K., & Huckauf, A. (2022). How sign language expertise can influence the effects of face masks on non-linguistic characteristics. *Cognitive Research: Principles and Implications*, *7*(1), 53. doi: 10.1186/s41235-022-00405-6.

Lau, W. K., & Huckauf, A. (2021). Effects of face masks on the appearance of emotional expressions and invariant characteristics. *Open Psychology*, *3*(1), 87–102. doi: 10.1515/psych-2020-0113.

Lau, W. K., Sauter, M., & Huckauf, A. (2022). Small Pupils Lead to Lower Judgements of a Person's Characteristics for Exaggerated, but Not for Realistic Pupils. *Behavioral Sciences*, *12*(8), 283. doi: 10.3390/bs12080283.

Lebrecht, S., Bar, M., Barrett, L., & Tarr, M. (2012). Micro-valences: Perceiving affective valence in everyday objects. *Frontiers in Psychology*, *3*(107), 1–5. doi: 10.3389/fpsyg.2012.00107.

LeDoux, J. E. (2012). Chapter 21 – Evolution of human emotion: A view through fear. In M. A. Hofman & D. Falk (Eds.), *Progress in brain research* (Vol. 195, pp. 431–442). Elsevier.

Lee, D. H., Susskind, J. M., & Anderson, A. K. (2013). Social transmission of the sensory benefits of eye widening in fear expressions. *Psychological Science*, *24*(6), 957–965. doi: 10.1177/0956797612464500.

Lyyra, P., Myllyneva, A., & Hietanen, J. K. (2018). Mentalizing eye contact with a face on a video: Gaze direction does not influence autonomic arousal. *Scandinavian Journal of Psychology*, *59*(4), 360–367.

Manesi, Z., Van Lange, P. A. M., & Pollet, T. V. (2016). Eyes wide open: Only eyes that pay attention promote prosocial behavior. *Evolutionary Psychology*, *14*(2), 1474704916640780. doi: 10.1177/1474704916640780.

Marsh, A. A., & Ambady, N. (2007). The influence of the fear facial expression on prosocial responding. *Cognition and Emotion*, *21*(2), 225–247. doi: 10.1080/02699930600652234.

Mathews, A., Fox, E., Yiend, J., & Calder, A. (2003). The face of fear: Effects of eye gaze and emotion on visual attention. *Visual Cognition*, *10*(7), 823–835. doi: 10.1080/13506280344000095.

Matsugasaki, K., Tsukamoto, W., & Ohtsubo, Y. (2015). Two failed replications of the watching eyes effect. *Letters on Evolutionary Behavioral Science*, *6*(2), 17–20. doi: 10.5178/lebs.2015.36.

Matsumoto, D. (1989). Face, culture, and judgments of anger and fear: Do the eyes have it? *Journal of Nonverbal Behavior*, *13*(3), 171–188. doi: 10.1007/BF00987048.

Mauchly, J. W. (1940). Significance test for sphericity of a normal n-variate distribution. *The Annals of Mathematical Statistics*, *11*(2), 204–209. doi: 10.1214/aoms/1177731915.

McCarthy, G., Puce, A., Gore, J. C., & Allison, T. (1997). Face-specific processing in the human fusiform gyrus. *Journal of Cognitive Neuroscience*, *9*(5), 605–610.

Midelfart, A. (1996). Women and men-same eyes? *Acta Ophthalmologica Scandinavica*, *74*(6), 589–592. doi: 10.1111/j.1600-0420.1996.tb00741.x.

Misailidi, P., & Bonoti, F. (2008). emotion in children's art: Do young children understand the emotions expressed in other children's drawings? *Journal of Early Childhood Research*, *6*(2), 189–200. doi: 10.1177/1476718x08088677.

Morris, J. S., deBonis, M., & Dolan, R. J. (2002). Human amygdala responses to fearful eyes. *NeuroImage*, *17*(1), 214–222. doi: 10.1006/nimg.2002.1220.

Myllyneva, A., & Hietanen, J. K. (2015). The dual nature of eye contact: To see and to be seen. *Social Cognitive and Affective Neuroscience*, *11*(7), 1089–1095. doi: 10.1093/scan/nsv075.

Nestor, A., & Tarr, M. J. (2008). Gender recognition of human faces using color. *Psychological Science*, *19*(12), 1242–1246. doi: 10.1111/j.1467-9280.2008.02232.x.

Nettle, D., Harper, Z., Kidson, A., Stone, R., Penton-Voak, I. S., & Bateson, M. (2013). The watching eyes effect in the Dictator Game: It's not how much you give, it's being seen to give something. *Evolution and Human Behavior*, *34*(1), 35–40. doi: 10.1016/j.evolhumbehav.2012.08.004.

Nettle, D., Nott, K., & Bateson, M. (2012). 'Cycle thieves, we are watching you': Impact of a simple signage intervention against bicycle theft. *PLOS ONE*, *7*(12), e51738. doi: 10.1371/journal.pone.0051738.

Nichols, K. A., & Champness, B. G. (1971). Eye gaze and the GSR. *Journal of Experimental Social Psychology*, *7*(6), 623–626. doi: 10.1016/0022-1031(71)90024-2.

Northover, S. B., Pedersen, W. C., Cohen, A. B., & Andrews, P. W. (2017). Effect of artificial surveillance cues on reported moral judgment: Experimental failures to replicate and two meta-analyses. *Evolution and Human Behavior*, *38*(5), 561–571. doi: 10.1016/j.evolhumbehav.2016.12.003.

Oda, R., & Ichihashi, R. (2016). The watching eyes effect on charitable donation is boosted by fewer people in the vicinity. *Letters on Evolutionary Behavioral Science*, *7*(2), 9–12. doi: 10.5178/lebs.2016.52.

Oda, R., Kato, Y., & Hiraishi, K. (2015). The watching-eye effect on prosocial lying. *Evolutionary Psychology*, *13*(3), 1474704915594959. doi: 10.1177/1474704915594959.

Olsson, A., Nearing, K. I., & Phelps, E. A. (2007). Learning fears by observing others: The neural systems of social fear transmission. *Social Cognitive and Affective Neuroscience*, *2*(1), 3–11. doi: 10.1093/scan/nsm005.

Panagopoulos, C. (2014). Watchful eyes: Implicit observability cues and voting. *Evolution and Human Behavior*, *35*(4), 279–284. doi: 10.1016/j.evolhumbehav.2014.02.008.

Park, J. H., Faulkner, J., & Schaller, M. (2003). Evolved disease-avoidance processes and contemporary anti-social behavior: Prejudicial attitudes and avoidance of people with physical disabilities. *Journal of Nonverbal Behavior*, *27*(2), 65–87. doi: 10.1023/A:1023910408854.

Pauwels, L., Declerck, C. H., & Boone, C. (2017). Watching eyes and living up to expectations: Unkind, not kind, eyes increase first mover cooperation in a sequential Prisoner's Dilemma. *Games*, *8*(2), 20. doi: 10.3390/g8020020.

Pfattheicher, S., & Keller, J. (2015). The watching eyes phenomenon: The role of a sense of being seen and public self-awareness. *European Journal of Social Psychology*, *45*(5), 560–566. doi: 10.1002/ejsp.2122.

Raihani, N. J., & Bshary, R. (2012). A positive effect of flowers rather than eye images in a large-scale, cross-cultural dictator game. *Proceedings of the Royal Society B: Biological Sciences*, *279*(1742), 3556–3564. doi: 10.1098/rspb.2012.0758.

Reed, L. I., & DeScioli, P. (2017). Watch out! How a fearful face adds credibility to warnings of danger. *Evolution and Human Behavior*, *38*(4), 490–495. doi: 10.1016/j.evolhumbehav.2017.03.003.

Rodin, M. J. (1987). Who is memorable to whom: A study of cognitive disregard. *Social Cognition*, *5*, 144–165. doi: 10.1521/soco.1987.5.2.144.

Rotella, A., Sparks, A. M., Mishra, S., & Barclay, P. (2021). No effect of 'watching eyes': An attempted replication and extension investigating individual differences. *PLOS ONE*, *16*(10), e0255531. doi: 10.1371/journal.pone.0255531.

Roy-Charland, A., Perron, M., Young, C., Boulard, J., & Chamberland, J. A. (2015). The confusion of fear and surprise: A developmental study of the perceptual-attentional limitation hypothesis using eye movements. *The Journal of Genetic Psychology*, *176*(5), 281–298. doi: 10.1080/00221325.2015.1066301.

Sacco, D. F., & Hugenberg, K. (2009). The look of fear and anger: Facial maturity modulates recognition of fearful and angry expressions. *Emotion*, *9*(1), 39. doi: 10.1037/a0014081.

Sato, W., & Yoshikawa, S. (2007). Enhanced experience of emotional arousal in response to dynamic facial expressions. *Journal of Nonverbal Behavior*, *31*(2), 119–135. doi: 10.1007/s10919-007-0025-7.

Schützwohl, A., & Reisenzein, R. (2012). Facial expressions in response to a highly surprising event exceeding the field of vision: A test of Darwin's theory of surprise. *Evolution and Human Behavior*, *33*(6), 657–664. doi: 10.1016/j.evolhumbehav.2012.04.003.

Sénémeaud, C., Sanrey, C., Callé, N., Plainfossé, C., Belhaire, A., & Georget, P. (2017). The watching-eyes phenomenon and blood donation: Does exposure to pictures of eyes increase blood donation by young adults? *Transfusion and Apheresis Science*, *56*(2), 168–170. doi: 10.1016/j.transci.2016.11.001.

Shen, W., & Cannella, A. A. (2002). Power dynamics within top management and their impacts on CEO dismissal followed by inside succession. *Academy of Management Journal*, *45*(6), 1195–1206. doi: 10.5465/3069434.

Shinohara, A., & Yamamoto, S. (2018). No evidence for the watching-eyes effect on human impulsivity. *Frontiers in Psychology*, *9*(1887), 1–7. doi: 10.3389/fpsyg.2018.01887.

Simion, F., Di Giorgio, E., Leo, I., & Bardi, L. (2011). The processing of social stimuli in early infancy: From faces to biological motion perception. *Progress in brain research*, *189*, 173–193.

Simion, F., & Giorgio, E. D. (2015). Face perception and processing in early infancy: Inborn predispositions and developmental changes. *Frontiers in Psychology*, *6*, 969.

Singer, G. G. (2010). Color in Ancient Egypt. *Terrae Antiqvae*. http://terraeantiqvae.com/profiles/blogs/color-in-ancient-egypt.

Sparks, A., & Barclay, P. (2015). No effect on condemnation of short or long exposure to eye images. *Letters on Evolutionary Behavioral Science*, *6*(2), 13–16. doi: 10.5178/lebs.2015.35.

Sueur, C., Piermattéo, A., & Pelé, M. (2023). Eye image effect in the context of pedestrian safety: A French questionnaire study [version 2; peer review: 2 approved]. *F1000Research*, *11*(218), 1–30. doi: 10.12688/f1000research.76062.1.

Tipples, J., Atkinson, A. P., & Young, A. W. (2002). The eyebrow frown: A salient social signal. *Emotion*, *2*(3), 288. doi: 10.1037/1528-3542.2.3.288.

Todorov, A., & Duchaine, B. (2008). Reading trustworthiness in faces without recognizing faces. *Cognitive Neuropsychology*, *25*(3), 395–410. doi: 10.1080/02643290802044996.

Todorov, A., Pakrashi, M., & Oosterhof, N. N. (2009). Evaluating faces on trustworthiness after minimal time exposure. *Social Cognition*, *27*(6), 813–833. doi: 10.1521/soco.2009.27.6.813.

Van Breen, J. A., De Dreu, C. K., & Kret, M. E. (2018). Pupil to pupil: The effect of a partner's pupil size on (dis) honest behavior. *Journal of Experimental Social Psychology*, *74*, 231–245. doi: 10.1016/j.jesp.2017.09.009.

Van Doesum, N. J., van der Wal, A. J., Boomsma, C., & Staats, H. (2021). Aesthetics and logistics in urban parks; can moving waste receptacles to park exits decrease littering? *Journal of Environmental Psychology*, *77*, 101669. doi: 10.1016/j.jenvp.2021.101669.

Vernon, L. L., & Berenbaum, H. (2002). Disgust and fear in response to spiders. *Cognition and Emotion*, *16*(6), 809–830. doi: 10.1080/02699930143000464.

Wang, C.-A., Baird, T., Huang, J., Coutinho, J. D., Brien, D. C., & Munoz, D. P. (2018). Arousal effects on pupil size, heart rate, and skin conductance in an emotional face task. *Frontiers in Neurology*, *9*, 1029.

Willis, J., & Todorov, A. (2006). First impressions: Making up your mind after a 100-ms exposure to a face. *Psychological Science*, *17*(7), 592–598. doi: 10.1111/j.1467-9280.2006.01750.x.

Yue, X., & Ng, S. H. (1999). Filial obligations and expectations in China: Current views from young and old people in Beijing. *Asian Journal of Social Psychology*, *2*(2), 215–226. doi: 10.1111/1467-839X.00035.

Zhang, Q. (2010). Asian Americans beyond the model minority stereotype: The nerdy and the left out. *Journal of International and Intercultural Communication*, *3*(1), 20–37. doi: 10.1080/17513050903428109.

Appendix

A1 Stimulus Selection Criteria

The eight models selected from the database were based on three criteria: 1) validation ratings of the emotional expression, 2) perceived age ratings, and 3) presence of obstruction around the eye region.

1) We took only the models who received the highest ratings in each perceived emotional expression. The rating data are available in the validation score spreadsheet from the face database.
2) We computed the average perceived age from the validation score spreadsheet. We then selected old models whose perceived age is >60 years old and young models whose perceived age <30 years old.
3) Of the remaining images, we ensured that the models did not have obstructions to the eyes and eye region (e.g., strands of hair partially blocking the eyes).

From these three criteria, we derived from the database model IDs for old faces: ID = 004, 146, 012, 060, and young faces: ID = 031, 057, 152, 182. The average perceived age of these images from the FACES database was 66.87 years old (SD = 3.36) for old faces and 26.44 years old (SD = 2.21) for young faces.

A2 List of Stereotypical Statements

Table A1 shows the stereotypical statements used in the experiment. The statements are grouped into the categories: race, gender, and religion. The statements are coded into positive and negative biases. Positive bias indicates a good stereotype, e.g., *"Meiner Meinung nach sind Frauen gut in Mathe" (In my opinion, women are good in Math)*. Negative bias indicates a bad stereotype, e.g., "*Meiner Erfahrung nach fotografieren Asiaten alles und jeden" (In my experience, Asians photograph everything and everyone)*.

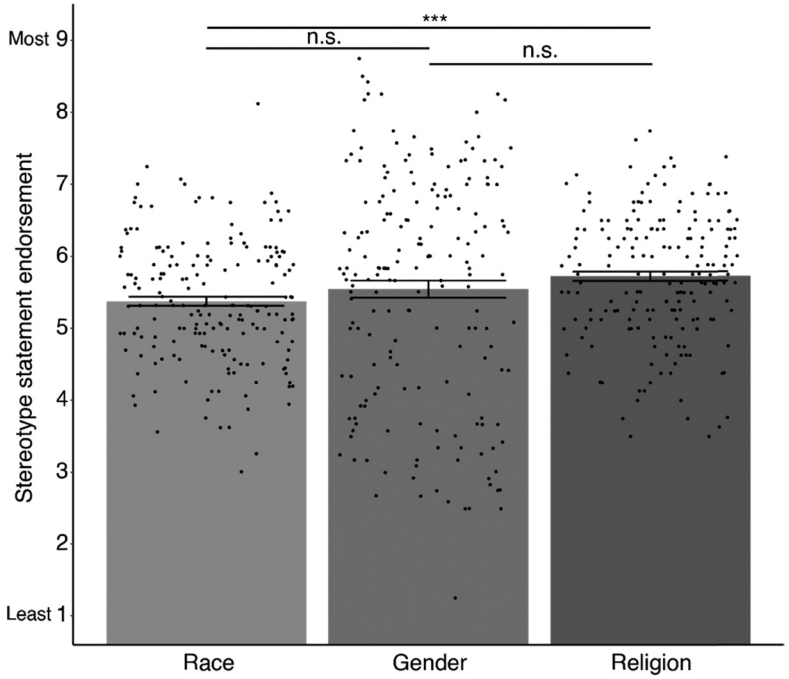

Figure A1: Bar graphs showing the stereotypical statement endorsement across race, gender, and religion stereotype categories. The scatter indicates responses from participants per category. Error bars illustrate standard error of means (SEM). *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$, n.s. – non-significant.

Table A1: Stereotypical statements used in the experiment with translations

#	Stereotype statement	Category	Coding
1	Meiner Meinung nach werden türkische Frauen unterdrückt *In my opinion, Turkish women are oppressed*	Race	Negative
2	Meiner Meinung nach sind türkische Männer friedlich *In my opinion, Turkish men are peaceful*		Positive
3	Meiner Meinung nach sind Türken Sozialschmarotzer *In my opinion, Turks are social parasites*		Negative
4	Meiner Meinung nach sind Türken gesetzestreu *In my opinion, Turks are law-abiding*		Positive
5	Meiner Erfahrung nach fotografieren Asiaten alles und jeden *In my experience, Asians take pictures of everything and everyone*		Negative
6	Meiner Erfahrung nach vertragen Asiaten reichlich Alkohol *In my experience, Asians can hold their liquor*		Positive
7	Meiner Meinung nach sind Asiaten gute Autofahrer *In my opinion, Asians are good drivers*		Positive
8	Meiner Meinung nach haben Asiaten strenge Eltern *In my opinion, Asians have strict parents*		Negative
9	Meiner Meinung nach denken Frauen anders *In my opinion, women think differently*	Gender	Negative
10	Meiner Meinung nach sind Frauen gut in Mathe *In my opinion, women are good at math*		Positive
11	Meiner Meinung nach sind Frauen hormongesteuert *In my opinion, women are hormone driven*		Negative
12	Meiner Meinung nach haben Frauen einen guten Orientierungssinn *In my opinion, women have a good sense of direction*		Positive
13	Meiner Meinung nach muss ein Mann nicht stark sein und seine Familie beschützen *In my opinion, a man doesn't have to be strong and protect his family*		Positive
14	Meiner Meinung nach sollten Männer niemals bereit sein, ihre Familie auch mit Gewalt zu verteidigen *In my opinion, men should never be willing to defend their family even with violence*		Positive
15	Meiner Meinung nach müssen Männer manchmal Härte zeigen *In my opinion, men sometimes need to show toughness*		Negative
16	Meiner Meinung nach können Männer besser mit Geld umgehen *In my opinion, men can handle money better*		Negative
17	Meiner Meinung nach tragen Juden die Verantwortung für die meisten Kriege auf der Welt *In my opinion, Jews are responsible for most of the wars in the world*	Race	Negative
18	Meiner Meinung nach reden Juden zu wenig über den Holocaust *In my opinion, Jews talk too little about the Holocaust*		Positive
19	Meiner Meinung nach haben Juden zu viel Macht in der Wirtschaft *In my opinion, Jews have too much power in the economy*		Negative
20	Meiner Meinung nach sind Juden politisch moderat *In my opinion, Jews are politically moderate*		Positive
21	Meiner Meinung nach vermindern Ausländer viele soziale Probleme *In my opinion, foreigners reduce many social problems*		Positive
22	Meiner Meinung nach verbessere sich meine Situation mit der Reduzierung der ausländischen Bevölkerung *In my opinion, my situation improves with the reduction of the foreign population*		Negative
23	Meiner Meinung nach gibt es zu wenige Ausländer *In my opinion, there are too few foreigners*		Positive
24	Meiner Meinung nach kommen die Ausländer nur hier her, um unseren Sozialstaat auszunutzen *In my opinion, foreigners come here only to take advantage of our welfare state*		Negative
25	Meiner Meinung nach passen Muslime in westliche Welt *In my opinion, Muslims fit into the Western world*	Religion	Positive
26	Meiner Meinung nach sind Muslime gefahrlos *In my opinion, Muslims are harmless*		Positive
27	Durch Muslime fühle ich mich wie Fremder im eigenen Land *Muslims make me feel like a stranger in my own country*		Negative
28	Meiner Meinung nach sollten Muslimen die Zuwanderung untersagt werden		Negative

(Continued)

Table A1: *Continued*

#	Stereotype statement	Category	Coding
	In my opinion, Muslims should not be allowed to immigrate		
29	Meiner Meinung nach sind Christen engstirnig		Negative
	In my opinion, Christians are narrow-minded		
30	Meiner Meinung nach ist das Christentum modern		Positive
	In my opinion, Christianity is modern		
31	Meiner Meinung nach sind Christen doppelmoralisch		Negative
	In my opinion, Christians are double-moral		
32	Meiner Meinung nach sind Christen tiefgründig		Positive
	In my opinion, Christians are profound		
33	Meiner Meinung nach ist Homosexualität unmoralisch	Gender	Negative
	In my opinion, homosexuality is immoral		
34	Meiner Meinung nach ist Homosexualität natürlich		Positive
	In my opinion, homosexuality is natural		
35	Meiner Meinung nach nimmt das Thema Homosexualität zu wenig Raum in den Medien ein		Positive
	In my opinion, the topic of homosexuality takes up too little space in the media		
36	Meiner Meinung nach sollten Homosexuelle aufhören, so einen Wirbel um ihre Sexualität zu machen		Negative
	In my opinion, homosexuals should stop making such a fuss about their sexuality		

A3 Comparison of Ratings Across Three Stereotype Categories

We checked whether participants could have responded to the three stereotype categories differently by running a 3 × 1 repeated measures ANOVA with the factor: stereotype categories (race, gender, and religion). The post hoc α value to evaluate significance for this comparison was 0.05 ÷ 3 = 0.0167.

There was a significant main effect of stereotype categories on stereotypical statement endorsement, F (1.37, 256.51) = 5.30, p = 0.013, η_p^2 = 0.028 (Figure A1). Post-hoc comparisons were conducted using Bonferroni correction: α = 0.0167. Endorsement differed between statements concerning *race* (M = 5.38, SD = 0.86) and *religion* (M = 5.72, SD = 0.89), t(187) = −4.73, p < 0.001, Cohen's d = −0.35, but not between *race* and *gender* (M = 5.54, SD = 1.62, p = 0.10), or between *gender* and *religion* (p = 0.19).

From the results, it is unlikely that the ratings across the entire experiment could be attributed to the differences between *race* and *religion*. This is because of the small effect size (Cohen's d = −0.35). Thus, subsequent analyses in the main manuscript excluded stereotype categories as a factor in simple data interpretation.

A4 Two-Way Interactions

We investigated the watching eyes effect on stereotypical statement endorsement by running a 4 × 2 × 2 repeated measures ANOVA. The main effects of the two-way interactions (emotional expressions × age, emotional expressions × sex) are reported here. Post-hoc comparisons were conducted for significant interaction effects.

We compared all possible emotional expression combinations, using an α value, 0.05 ÷ 6 = 0.008. For eye age × emotional expressions, we computed the differences of each emotional expression by eye age, using an α value, 0.05 ÷ 4 = 0.0125. For eye sex × emotional expression, we determined the differences of each emotional expression by eye sex, using an α value, 0.05 ÷ 4 = 0.0125.

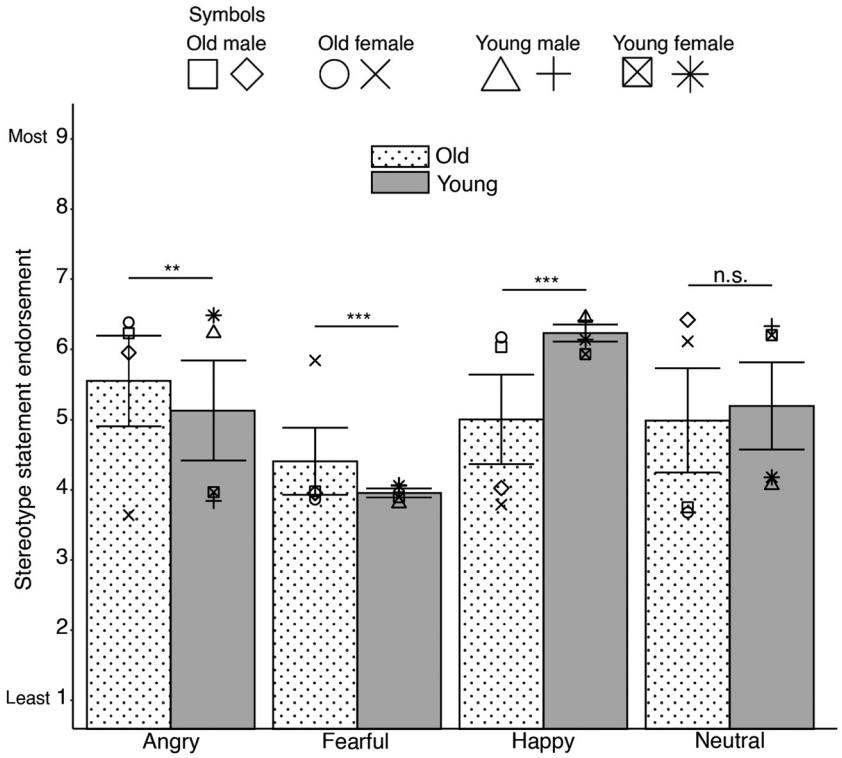

Figure A2: Bar graphs showing the two-way interactions between emotional expression and age. Symbols indicate mean responses for each eye model. Error bars illustrate the standard error of means (SEM). *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$, and n.s. – non-significant.

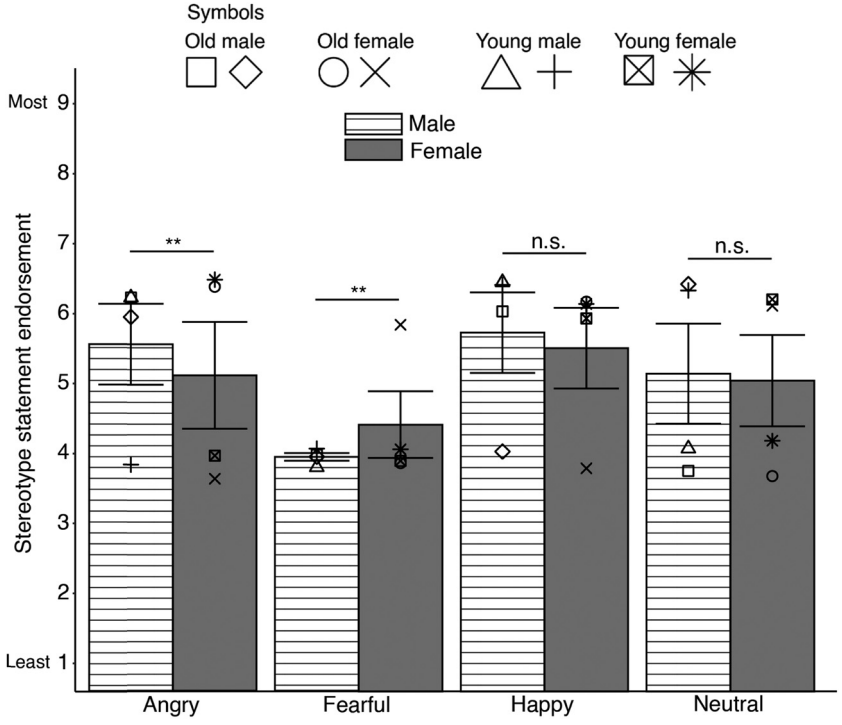

Figure A3: Bar graphs showing the two-way interactions between emotional expression and sex. Symbols indicate mean responses for each eye model. Error bars illustrate the standard error of means (SEM). ***$p < 0.001$, **$p < 0.01$, *$p < 0.05$, and n.s. – non-significant.

Table A2: Summary of factors and variables which violated sphericity assumptions

Factor	Mauchly's W	χ^2	Violation?
Stereotype categories	0.54	113.94***	✓
Emotional expression (EE)	0.88	24.66***	✓
EE × age	0.99	1.30	
EE × sex	0.98	3.48	
EE × age × sex	0.98	3.94	

***$p < 0.001$.

A4.1 Interaction Between Emotional Expressions and Age Influence the Stereotypical Statement Endorsement

There was a significant interaction between emotional expressions and age, $F(3, 561) = 33.38$, $p < 0.001$, and $\eta_p^2 = 0.15$ (Figure A2). Post hoc comparisons using Bonferroni-corrected $\alpha = 0.0125$ were conducted. Participants show higher endorsement when seeing old angry eyes ($M = 5.55$, SD = 1.40) than young angry eyes ($M = 5.13$, SD = 1.22), $t(187) = 3.03$, $p = 0.003$, and Cohen's $d = 0.22$, and when seeing old fearful eyes ($M = 4.41$, SD = 1.38) than young fearful eyes ($M = 3.96$, SD = 1.56), $t(187) = 3.31$, $p = 0.001$, and Cohen's $d = 0.24$.

Stereotype endorsements were lower for old happy eyes ($M = 5.00$, SD = 1.30) than young happy eyes ($M = 6.23$, SD = 1.45), $t(187) = -9.20$, $p < 0.001$, and Cohen's $d = -0.67$. There were no differences for neutral old eyes ($M = 4.99$, SD = 1.38) and neutral young eyes ($M = 5.20$, SD = 1.33), $t(187) = -1.51$, $p = 0.13$, and Cohen's $d = -0.11$.

A4.2 Interaction between emotional expressions and sex influence stereotypical statement endorsement

There was a significant interaction between emotional expressions and sex, $F(3, 561) = 8.01$, $p < 0.001$, $\eta_p^2 = 0.04$ (Figure A3). Post-hoc comparisons using Bonferroni-corrected $\alpha = 0.0125$ were conducted. The extent to endorse stereotype statements was higher when seeing male angry eyes ($M = 5.56$, SD = 1.33) than female angry eyes ($M = 5.12$, SD = 1.35), $t(187) = 3.10$, $p = 0.002$, and Cohen's $d = 0.23$. Stereotype statement endorsement were lower for male fearful eyes ($M = 3.95$, SD = 1.53) than for female fearful eyes ($M = 4.41$, SD = 1.50), $t(187) = -3.19$, $p = 0.002$, and Cohen's $d = -0.23$.

There were no differences between male happy eyes ($M = 5.73$, SD = 1.34) and female happy eyes ($M = 5.51$, SD = 1.36), $t(187) = 1.74$, $p = 0.08$, and Cohen's $d = 0.13$. There were no differences between male neutral eyes ($M = 5.14$, SD = 1.38) and female neutral eyes ($M = 5.04$, SD = 1.32), $t(187) = 0.74$, $p = 0.46$, and Cohen's $d = 0.05$.

A5 Statistical Assumption Checks for Repeated Measures ANOVA

Statistical assumption checks were conducted for the repeated measures ANOVA. We checked whether the sphericity assumption was violated using Mauchly's W (Mauchly, 1940). Greenhouse–Geisser correction was applied when comparisons violated the sphericity assumption (Greenhouse & Geisser, 1959). The results of the statistical checks are reported in Table A2.