

Article

Small Pupils Lead to Lower Judgements of a Person's Characteristics for Exaggerated, but Not for Realistic Pupils

Wee Kiat Lau , Marian Sauter  and Anke Huckauf

General Psychology, Institute of Psychology and Pedagogics, Ulm University, Albert-Einstein-Allee 47, 89081 Ulm, Germany

* Correspondence: wee.lau@uni-ulm.de

Abstract: Our eyes convey information about a person. The pupils may provide information regarding our emotional states when presented along with different emotional expressions. We examined the effects of pupil size and vergence on inferring other people's characteristics in neutral expression eyes. Pupil sizes were manipulated by overlaying black disks onto the pupils of the original eye images. The disk area was then changed to create small, medium, and large pupils. Vergence was simulated by shifting the medium-sized disks nasally in one eye. Pupil sizes were exaggerated for Experiment 1 and followed values from the literature for Experiment 2. The first Purkinje image from the eye photos in Experiment 2 was kept to preserve image realism. The characteristics measured were sex, age, attractiveness, trustworthiness, intelligence, valence, and arousal. Participants completed one of two online experiments and rated eight eye pictures with differently sized pupils and with vergence eyes. Both experiments were identical except for the stimuli designs. Results from Experiment 1 revealed rating differences between pupil sizes for all characteristics except sex, age, and arousal. Specifically, eyes with extremely small pupil sizes and artificial vergence received the lowest ratings compared to medium and large pupil sizes. Results from Experiment 2 only indicated weak effects of pupil size and vergence, particularly for intelligence ratings. We conclude that the pupils can influence how characteristics of another person are perceived and may be regarded as important social signals in subconscious social interaction processes. However, the effects may be rather small for neutral expressions.

Keywords: pupils; person; characteristics; eyes; uncanny valley



Citation: Lau, W.K.; Sauter, M.; Huckauf, A. Small Pupils Lead to Lower Judgements of a Person's Characteristics for Exaggerated, but Not for Realistic Pupils. *Behav. Sci.* **2022**, *12*, 283. <https://doi.org/10.3390/bs12080283>

Received: 11 July 2022

Accepted: 10 August 2022

Published: 12 August 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

There is a popular saying that the eyes are the gateway to a person's soul. Humans are born with a preference of looking at the eyes. Newborns spend a significantly longer time looking at a face with opened eyes than the same face with closed eyes [1]. The preference to look at faces with opened eyes continues to be higher than the preference to look at faces with closed eyes as the child matures [2]. We inevitably look at a person's eyes during social interaction. Sometimes, we steal quick glances at a stranger's eyes on the streets. We even make prolonged eye contact with intimate partners and loved ones. Therefore, the eyes are an important aspect of our daily habits.

First impressions are the characteristics we perceive about another person by looking at them [3–5]. We also perceive a person's characteristics by looking at their eyes and eye region, which is the area surrounding the eyes (between the temples and the nose bridge, including the eyebrows). We may judge a person's sex by looking at their eyes, as females tend to have smaller eyes than males [6]. The eyebrows can quickly change the perceived emotion of the eyes. For example, V-shaped brows appear angry and ^-shaped brows look sad [7]. We perceive a smile as more genuine when we see wrinkles (i.e., Crow's feet) surrounding the eye's edges [8]. These wrinkles may also make the eyes look older [9]. Eyebags develop underneath the eyes when a person is sleep deprived. They are often

perceived as tiredness [10]. Overall, we perceive some characteristics about a person when looking at their eye region.

Some studies also suggest that we can tell what a person is thinking by looking at where the eyes are gazing. In the Reading the Mind in the Eye's (RMET) test, characteristics about a person can be perceived by where the eyes are looking [11]. In the experiment, participants picked an adjective to describe what the pairs of black and white eyes were conveying. Participants were generally good at associating what the eyes conveyed when looking at the gaze. Hence, a person's characteristics are perceived when looking at the eye gaze.

1.1. Perceiving Characteristics about a Person from the Pupils and Vergence

The eyes convey much information about a person, besides the gaze alone. Some studies allude to the idea that the characteristics of another person can be perceived by looking at the pupils. These may be achieved by varying either pupil sizes, or pupil positions (i.e., vergence). Varying pupil sizes can influence how the emotional state of a person is perceived. A sad face appeared much sadder for constricted pupils than dilated pupils [12]. Pupil sizes can also influence the intensity of perceived fear [13].

Pupil sizes may also change the perception of abstract characteristics like attractiveness and trustworthiness. For instance, large pupils and large eyes are perceived as more attractive [14,15]. Neutral eyes appear more trustworthy when the pupils are dilated, and less trustworthy when the pupils are constricted [16,17]. These findings are replicated regardless of whether participants saw Caucasian or Asian eyes with neutral expressions.

Eye vergence tells us at which distance a person is fixating. Our eyes rotate nasally inward to focus on a very near object [18,19]. This contributes to the feeling of being looked at which itself suggests close and friendly relationships [20]. Sometimes, we appear cross-eyed if the object is extremely near to us. According to the literature, looking slightly cross-eyed could also influence how we are perceived by others. Individuals with convergence strabismus exhibit symptoms where one or both pupils are constantly turned nasally inward. As a result, they appear less intelligent due to their eye vergence [21].

These studies show that changes to pupil sizes and vergence may influence how specific characteristics are read from the eyes. Some of these studies also show that changes to pupil sizes and vergence, in relation to the full face, influences how we read certain characteristics about a person. It is currently unknown how much this effect can be elicited solely by the eyes and eye region alone. It is also unknown if changes to one aspect of the pupils can influence how several characteristics are read. For instance, do changes in pupil sizes influence perceived attractiveness, trustworthiness, and intelligence?

1.2. Study Goal

The goal of the study was to explore the role of the pupils in reading characteristics of a person by looking at the person's eyes. Previous studies adopted a mixture of full faces and eyes. For our study, we chose to use only the eyes such that other cues from the face would not influence the perceived impressions. Studies that used eye stimuli also tested dynamic pupil changes [20,21]. Since motion captures attention [22], it might be difficult to tell whether participants relied on the motion of dilating/constricting pupils or the pupils themselves to evaluate a person's characteristics. We circumvent this issue by showing participants only the pupils' end state (i.e., after dilation/constriction). Thus, we presented the following pupils in the study: big, medium, small, and vergence.

There were two experiments to the study. Both experiments were identical except for the stimuli shown. In each experiment, the pupils were manipulated differently. In Experiment 1, we presented pairs of eyes with black disks substituting the pupils. In doing so, we exaggerated pupil sizes differences, to make sure that pupil size differences can be easily perceived by the observers instantly. Vergence was also exaggerated by shifting medium-sized pupils. In Experiment 2, pupil sizes were varied according to real variations. In addition, the inserted pupils included reflections (i.e., first Purkinje image).

According to the literature, the eyes and eye region can convey information about a person's sex and age. Differently sized pupils can influence perceived attractiveness and trustworthiness and the emotional state of a person. Vergence can also influence how we perceived intelligence from another individual. Therefore, we measured following perceived characteristics in both experiments: sex, age, attractiveness, trustworthiness, intelligence, realism, familiarity, arousal, and valence. One's emotional state is commonly measured using arousal and valence as indicators of emotions from two dimensions of the Self-Assessment Manikin (SAM) scale [23]. We added realism and familiarity as additional variables to rule out whether the ratings could be due to memory effects if participants had previously seen the stimuli.

We hypothesized that pupil sizes (large, medium, and small) would influence the ratings for attractiveness, trustworthiness, intelligence, arousal, and valence based on previous literature. Specifically, large pupils should be rated higher in these characteristics than the other pupil sizes, and small pupils would receive the lowest scores [12,17,24]. We expected that vergence pupils (i.e., crossed-eye) would result in low ratings for intelligence than straight looking pairs of eyes [21]. Sex and age served as kind of control, since eye size and wrinkles affect the perceived sex and age of the eyes, respectively [6,9]. Finally, realism ratings should be higher in Experiment 2 using the real variations of pupils with reflections.

2. Methods

2.1. Participants

All participants were recruited in accordance with the Declaration of Helsinki and the study was approved by the ethics committee of Ulm University within the context of the SPP2199 project (see funding). Respondents provided informed consent prior before taking part in the study. In both experiments, 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. A power analysis yields a target sample size, $n = 71$, for our critical hypotheses t-tests, with the standard 0.8 power and alpha of 0.05.

265 participants participated in Experiment 1 between 24 Mar 2021 and 30 May 2021. This sample was composed mainly of young adults (Mean age = 29.59 years old, $SD = 13.06$, range = 18–76). Most were female ($n = 180$, 68.20%), followed by male ($n = 83$, 31.40%), and diverse ($n = 1$, 0.40%). A new group of 110 respondents participated in Experiment 2 between 07 Sep 2021 and 31 Dec 2021. The sample was composed mainly of young adults (Mean age = 23.25 years old, $SD = 7.04$, range = 18–54 years old). Most were female ($n = 91$, 82.70%), followed by male ($n = 16$, 14.5%), and diverse ($n = 3$, 2.70%).

2.2. Stimuli

The stimuli were pairs of eyes taken from the FACES database [25]. This database features colored Caucasian faces of young, middle-aged, and old models. Images from the database are set in portrait with the dimensions: 2835×3543 pixels, 300 ppi, 8 bits depth. Eight neutral face images (two young females, two old females, two young males, two old males) were selected from the database for the experiments. These faces were used in a previous work, and we reused them for the following experiments [26].

Image manipulation was carried out using GIMP (version 2.10). The eye stimuli for the present study were created by cropping the eyes from the database face images (Figure 1a). Eye region features like eyebrows were excluded. The resulting dimensions for the eyes were 1140×226 pixels. The pupils of each eye stimulus were manipulated to create large, medium, small sizes. The general idea was to cover the pupils by overlaying them with disks. Thereafter, the disk's area was adjusted to form different sizes (big, medium, small). We also created simulated vergence based on medium pupil size to simplify the complexity of the study. The idea was to shift the medium pupil size disks nasally to form the simulated vergence. Shifting the medium pupil size disks did not change the disk's area. Only the

disk's horizontal position was changed. Thus, we created a total of four (large, medium, small, vergence) images times eight eye models, (32 eye stimuli in total).

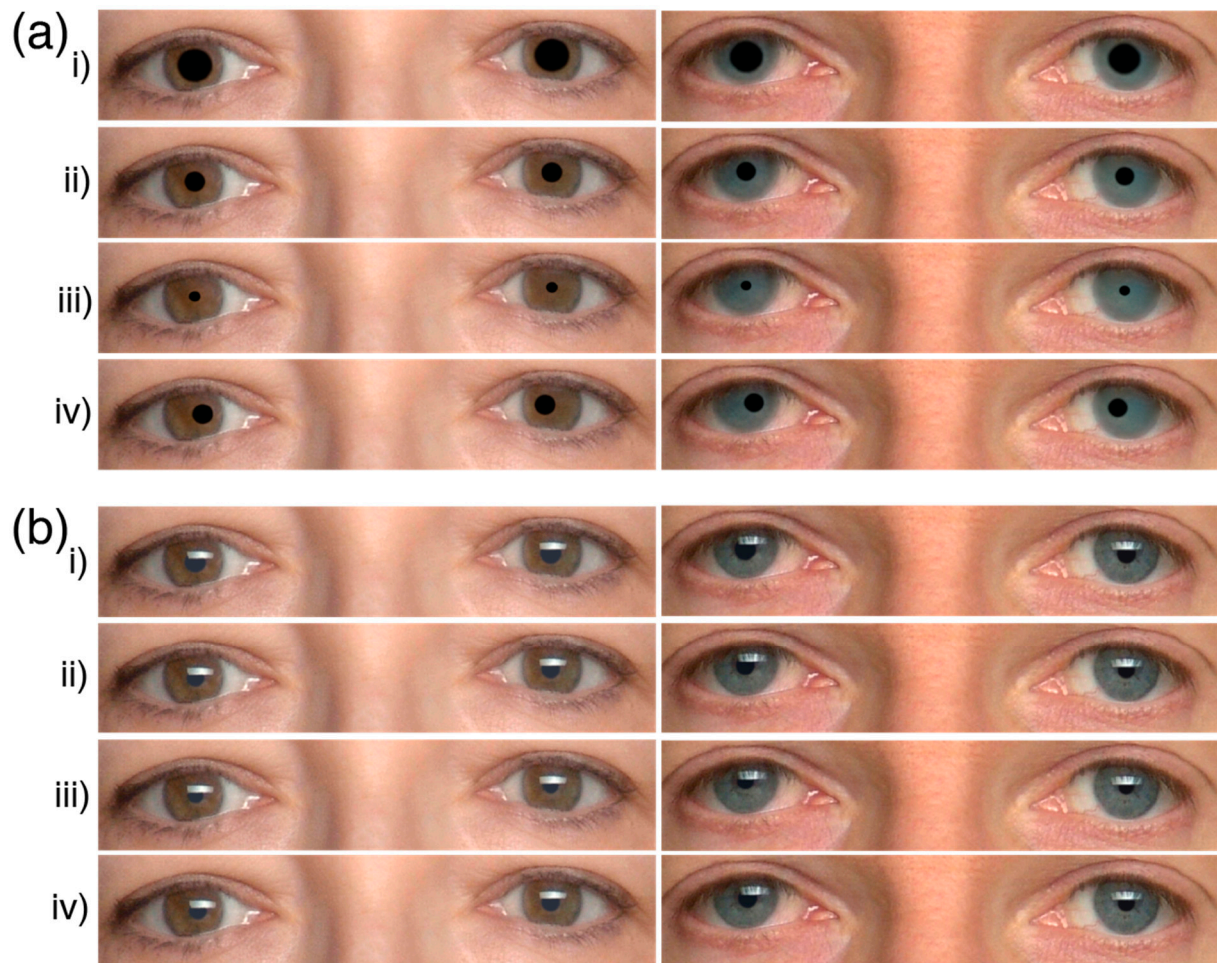


Figure 1. Illustrations of the stimuli used in (a) Experiment 1 and (b) Experiment 2. The eyes were 1140×226 pixels in dimensions. The pupils were (i) big, (ii) medium, (iii) small, and (iv) vergence. The pupils in (a) have the same area across all eye models. The pupils in (b) have different areas based on the eye models. The models in this illustration were models #29 and #70 from the FACES database.

In Experiment 1, all models have the same black disks in both eyes (see Figure 1a). The black disks had the following dimensions: large pupils 78×74 pixels, medium pupils 46×44 pixels, and small pupils 26×24 pixels. Simulated vergence was derived by displacing the left medium pupil 16 pixels and the right medium pupil 14 pixels nasally (i.e., horizontally towards the nose). The pupils were not rotated inwards equally since each eye rotates asymmetrically during vergence movements [27,28].

In Experiment 2, the pupil sizes were manipulated independent of the eye model according to previous literature [17]. In particular, the pupil sizes in Harrison, Wilson [12] ranged 60–167% and those in Kret, Fischer [16] ranged 60–140%. We used the range 50–150%. Like in Experiment 1, we first created black disks for the eye. Each eye model has a different pupil size. From here, the disk's area was then changed based on the following: large pupils +50% area, medium pupils +0% area, small pupils –50% area. The eye makes small rotations when looking at something far [28]. Therefore, we simulated the vergence of looking at far objects by displacing the left medium pupil five pixels nasally. The camera glare was included to preserve the realism of the original photos and to preserve the first Purkinje image (see Figure 1b).

2.3. Design

The experimental designs for both experiments were modifications of a previous work [26]. Experiments 1 and 2 were identical except for the stimuli shown. Participants finished only one experiment. In either experiment, participants were tasked to complete the same online survey. The online survey was administered using the EFS Survey by Questback GmbH [29]. The survey format, how it looks, and the questions in each section were the same in both experiments.

There were four survey versions (Figure 2a) since the 32 stimuli were distributed using a Latin-square method. The purpose was to ensure that participants never saw the same eye model again with different pupils, as prior memory could influence responses when seeing the same eye model. There were four sections per survey. The sections were always presented in the same order (Figure 2b). In Section 1, participants reported their arousal and valence. Arousal and valence were measured using the Self-Assessment Manikin (SAM) scale [23]. This section was repeated at the end of the survey. In Section 2, participants rated perceived sex, age, attractiveness, trustworthiness, intelligence, realism, and familiarity. In Section 3, participants rated the perceived emotional state of the stimuli (arousal, valence) using the SAM scale. In Section 4, participants completed questions concerning demographics (sex, age, education).

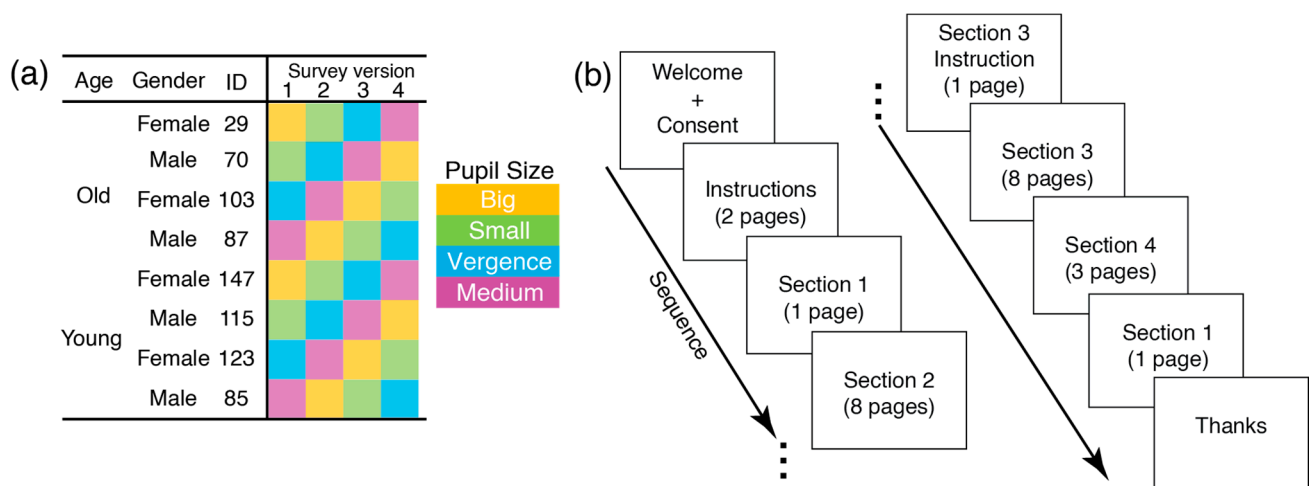


Figure 2. Survey version and sequence for both experiments. (a) illustrates the survey versions based on the Latin-square for presenting the stimuli. There were eight eye models. The age, gender, and ID of the models were counterbalanced. Each eye model never appeared more than once within a survey version. The colors correspond to various pupils. (b) shows the survey sequence. All participants completed the same sequence.

The eyes were shown in Sections 2 and 3. Each pair of eyes appeared at the top of the screen. The eyes appeared in random order. The eye order was different in Sections 2 and 3. Participants evaluated the eyes and provided their responses by selecting a dial or by sliding a point along the Likert scale (Figure 3). All scales used in the survey were based on the nine-point Likert scale, except when answering the demographics. The Likert scale's left pole indicated the least ("not at all . . .") and right pole indicated the most ("very . . .") for the following characteristics: attractiveness, trustworthiness, intelligence, realism, and familiarity. The sex scale was labelled very masculine on the left pole and very feminine on the right pole. The age scale was labeled in 5-year increments, ranging 20–24, 25–29, . . . , >60 years old. The arousal SAM scale was low arousal on the left, and high arousal on the right. The valence SAM scale was negative valence on the left, and positive valence on the right.

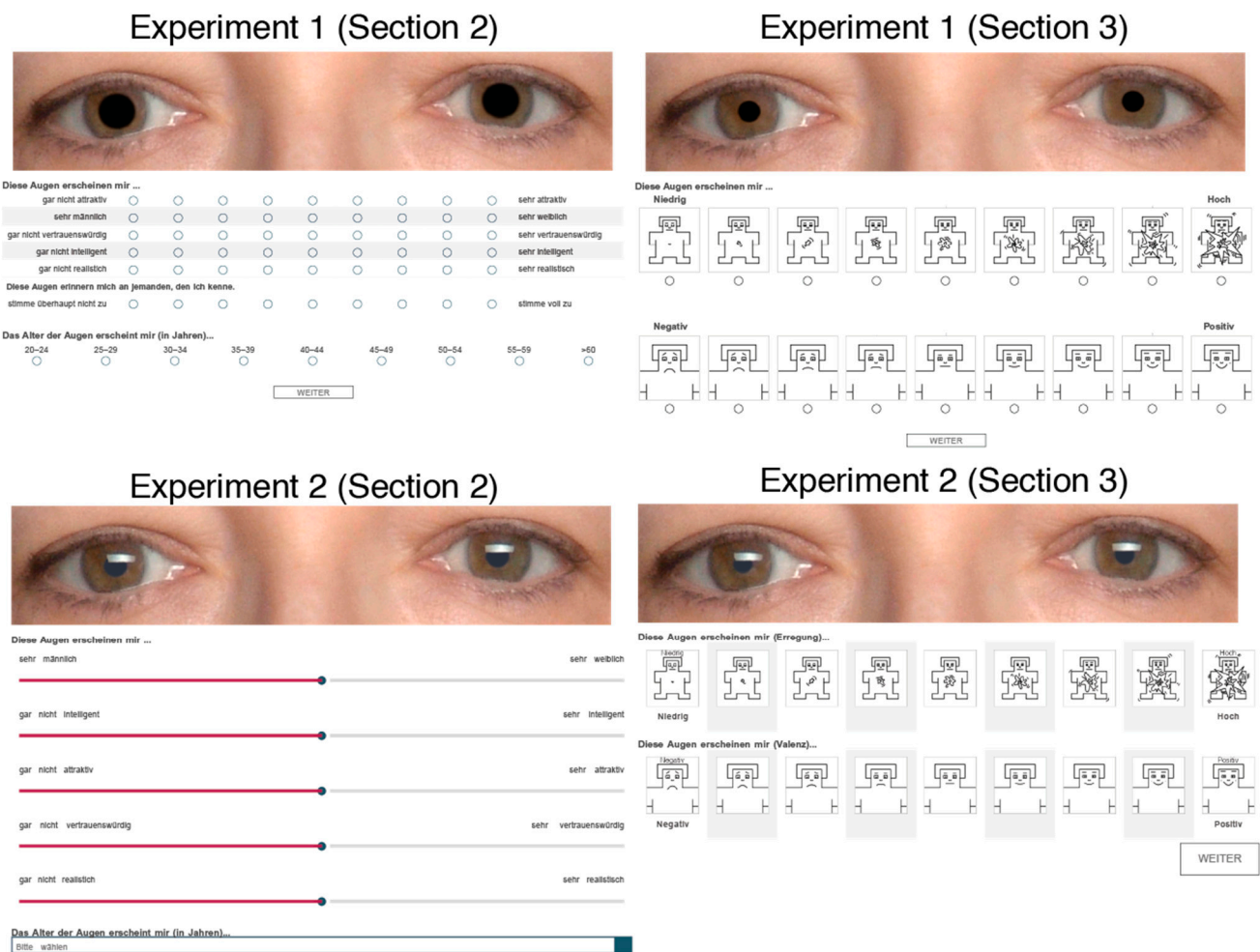


Figure 3. Sample screenshots of the survey layout in Sections 2 and 3 of the survey when the survey is not completed on a mobile device (i.e., mobile phone). In both experiments, the stimulus always appeared at the top of the screen. For illustration purposes, the eyes in Section 2 depict large pupils and those in Section 3 depict medium pupils.

We implemented two attention check questions in Section 1 to ensure that participants understood the SAM scale. Participants must pick the correct adjectives describing high arousal and positive valence. The adjectives for arousal and valence are described later in the manuscript. Participants must match the correct option on the Likert-scale.

The stimulus always appeared at the top of the screen (Figure 3). The question “This pair of eyes seem to me . . . ” was positioned directly below the stimulus, left-aligned to the survey page. The items to be rated were presented center-aligned within the page. Participants who completed the survey on a mobile device were encouraged to use the landscape orientation on their devices for the survey. Participants navigated to the next survey page by clicking a button at the bottom of the page. There was no backward button and participants could not return to previous pages to change their responses. Participants had to answer all of the questions. The message “please answer this question” appeared when participants failed to provide a response.

Participants responded to eight stimuli. Each of the eight stimuli was presented once in Section 2 (sex, age, attractiveness, trustworthiness, intelligence, realism, and familiarity) and once in Section 3 (arousal and valence). The stimuli that appeared in Sections 2 and 3 were the same, except the stimuli order in each section was randomized. Participants rated the stimuli across 16 pages in the survey. There were three instructions pages in the survey. There were 2 pages for Section 1, and 3 pages for Section 4 (demographics).

In total, participants navigated through 24 pages of the survey. The entire survey took approximately 10–20 min.

2.4. Procedure

The survey was administered online without extra supervision. Participants took part in the survey by clicking on a weblink. The link was sent through the Ulm University student mailing list and posted on social media (i.e., Facebook and Instagram). The survey was conducted in the German language. Participants saw the welcome instructional screen upon clicking on the survey link. The welcome screen indicated that the survey would last approximately 10–20 min. Participants were told that there were no right or wrong responses for the survey, and that the demographic information collected could not be used to identify them. Informed consent was given by clicking on the “I agree” button to begin the survey.

A new instruction screen was shown. Participants were briefed about the number of pages in the survey and the types of scales they would encounter. Participants navigated to the next survey page by clicking on a button at the bottom of the page. The instruction screen for Section 1 was presented. Participants were introduced to the pictographic SAM scale, along with the adjectives which described arousal and valence. These adjectives were derived from previous literature [30]. The adjectives for low arousal were relaxed, calm, sluggish, clumsy, sleepy, and rested. The adjectives for high arousal were stimulated, excited, turbulent, nervous, awake, and restless. The adjectives for negative valence were worried, annoyed, dissatisfied, moody, sad, desperate, and bored. The adjectives for positive valence were happy, content, satisfied, comfortable, hopeful, and relaxed. Participants completed two attention check questions to ensure that they comprehended the descriptions for arousal and valence scales. Participants were then randomly assigned to one of four survey versions.

The survey started with Section 1 (Figure 2b). Participants reported their current arousal and valence by choosing an item on the nine-point Likert scale. Section 2 began immediately. Participants saw a pair of eyes and several characteristics beneath it. Participants evaluated the stimulus based on the characteristics before proceeding to the next page. In Section 2, participants rated eight pairs of eyes regarding the perceived sex, age, attractiveness, trustworthiness, intelligence, realism, and familiarity. After the eighth stimulus was evaluated, participants entered Section 3. An instruction screen was presented to remind the participants the adjectives describing arousal and valence. Participants then evaluated the perceived arousal and valence on the same pairs of eyes they saw before, but in a different order. After evaluating the last stimulus, participants entered Section 4 to report their demographics. Section 1 was presented again to measure the post-survey arousal and valence. Finally, a thank you screen was shown to thank the participants for their contribution.

Some variables are now clarified to avoid confusion for the remaining manuscript. The SAM scale was used to measure participants’ arousal and valence (Section 1) and the perceived emotional state of the stimuli (Section 3). Participants’ arousal and valence were not analyzed in this manuscript since it was unrelated to the study goal. Therefore, arousal and valence refer to the participants’ ratings when participants evaluated the eyes.

2.5. Analysis

The analyses from both experiments were conducted in SPSS (IBM SPSS Statistics for Windows, Version 26.0). The dependent variables (DVs) and independent variables (IV) for both experiments were the same. The DVs were scores for each characteristic of sex, age, attractiveness, trustworthiness, intelligence, realism, familiarity, arousal, and valence. The IV was pupil size (large, medium, small) and vergence.

Data inclusion was run for Experiments 1 and 2 prior to other analyses. The goal was to analyze data from participants who did not fail the attention-check questions. In Experiment 1, data from 29 (10.94%) participants were not included since they selected the wrong items for the attention-check questions. Thus, we analyzed data of 236 (89.06%)

participants for Experiment 1. In Experiment 2, data from three (2.73%) participants were rejected since they answered the attention-check question incorrectly. The remaining participants ($n = 107$, 97.27%) were included for the analysis.

Several variables were re-coded. The critical variable for comprehending this manuscript is the DV, sex. The Likert scale for sex ranged between very masculine to very feminine. This variable was re-coded for the analysis such that higher values indicate greater accuracy at identifying the sex of the eye models.

The goal of the study was to determine effect sizes of pupil size and vergence on the estimation of various characteristics. Both experiments were analyzed the same way following previous studies [26,31]. We evaluated the influence of the pupils on perceived characteristics by running a 4×1 repeated measures ANOVA for each DV. We conducted post-hoc pairwise comparisons for DVs with significant main effects to further investigate the specific contributions of the pupils on the ratings. A conservative approach to account for inflated α values was adopted. The significance value for the post-hoc tests was determined using Bonferroni-correction: $0.05 \div 6 = 0.008$ across the four pupil sizes [32].

Statistical assumption violations were tested before running all analyses. Sphericity assumptions and appropriate corrections were used for the repeated measures ANOVA. We determined whether sphericity assumption was violated using Mauchly's W [33]. Greenhouse-Geisser correction was applied when comparisons violated the sphericity assumption [34]. Results concerning the statistical assumption tests were reported in Appendix A to streamline the readability of the manuscript.

We ran an additional correlational analysis between the attractiveness, trustworthiness, and intelligence ratings since these ratings are often associated with each other [35,36]. We did not find any consistencies in correlations across these variables in both experiments. Results of the correlations for each experiment were reported in Appendix B.

3. Results

3.1. Pupils Do Not Influence Perceived Sex and Age

Figure 4a shows the ratings for perceived sex and age in Experiment 1. There were no differences for sex ratings, $F(2.82, 662.01) = 0.31$, $p = 0.81$, $\eta_p^2 = 0.001$, and age ratings, $F(3, 705) = 1.21$, $p = 0.31$, $\eta_p^2 = 0.005$. Figure 4b illustrates the perceived sex and age ratings in Experiment 2. We found no differences for sex ratings, $F(2.74, 290.12) = 0.54$, $p = 0.66$, $\eta_p^2 = 0.005$ and age ratings, $F(3, 705) = 0.11$, $p = 0.95$, $\eta_p^2 = 0.001$.

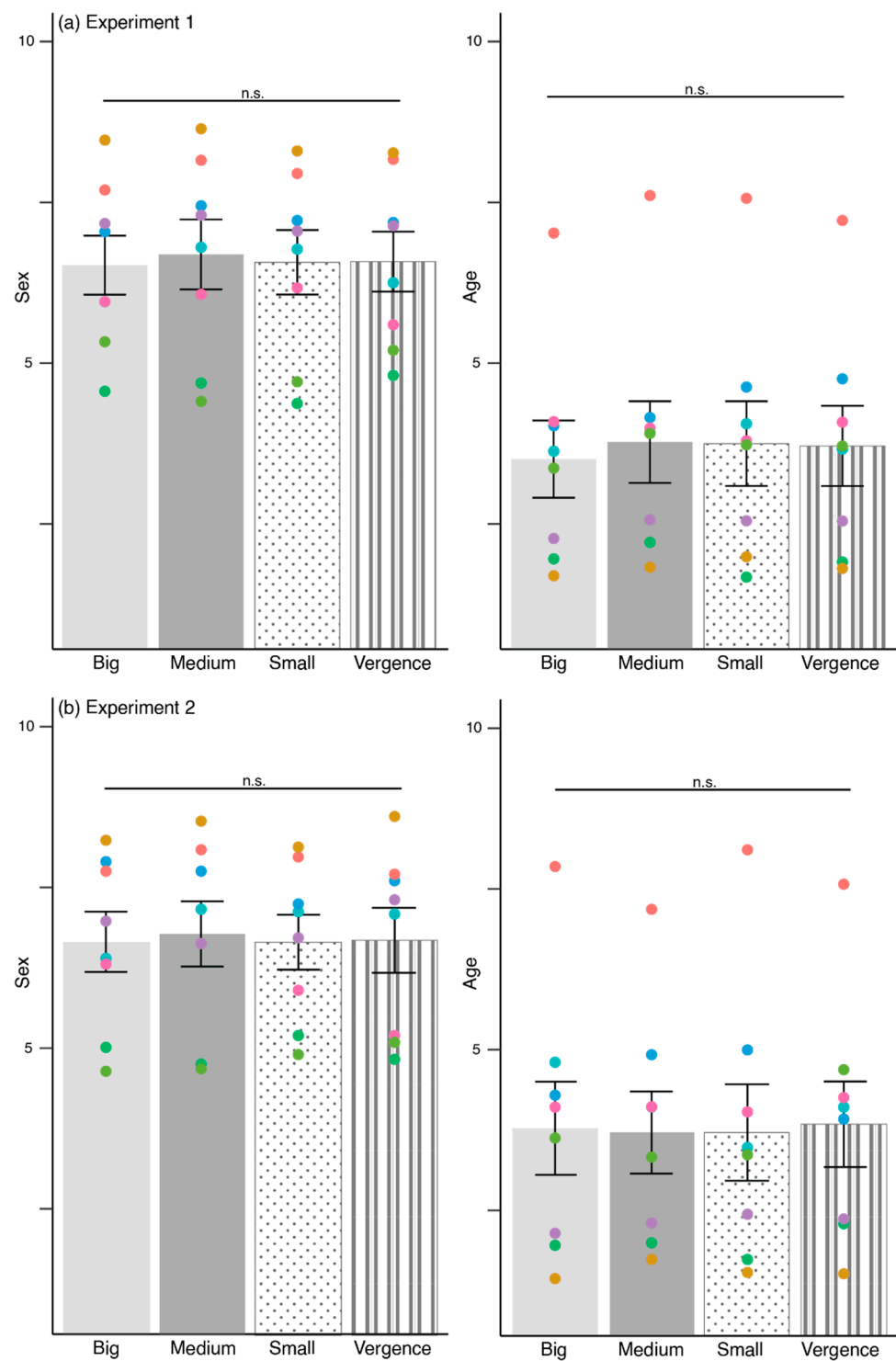


Figure 4. Sex and age ratings in (a) Experiment 1, and (b) Experiment 2 across big, medium, small, and vergence pupils. The colored scatter corresponds to each eye model. Error bars depict standard error of means, S.E.M. n.s.—non-significant.

In summary, the pupils did not influence perceived sex and age ratings.

3.2. Unrealistic Pupils Influence Perceived Attractiveness, Trustworthiness, and Intelligence

Figure 5a visualizes the ratings of attractiveness, trustworthiness, and intelligence in Experiment 1. The main effect of attractiveness ratings reached significance, $F(2.77,$

650.38) = 36.85, $p < 0.001$, $\eta_p^2 = 0.14$. Post-hoc pairwise comparison was conducted for attractiveness ratings using Bonferroni-correction ($p < 0.008$) and the results were presented in Table 1. The comparisons showed that both big and medium pupils appeared more attractive than small pupils and vergence. Small pupils also looked more attractive than vergence ones.

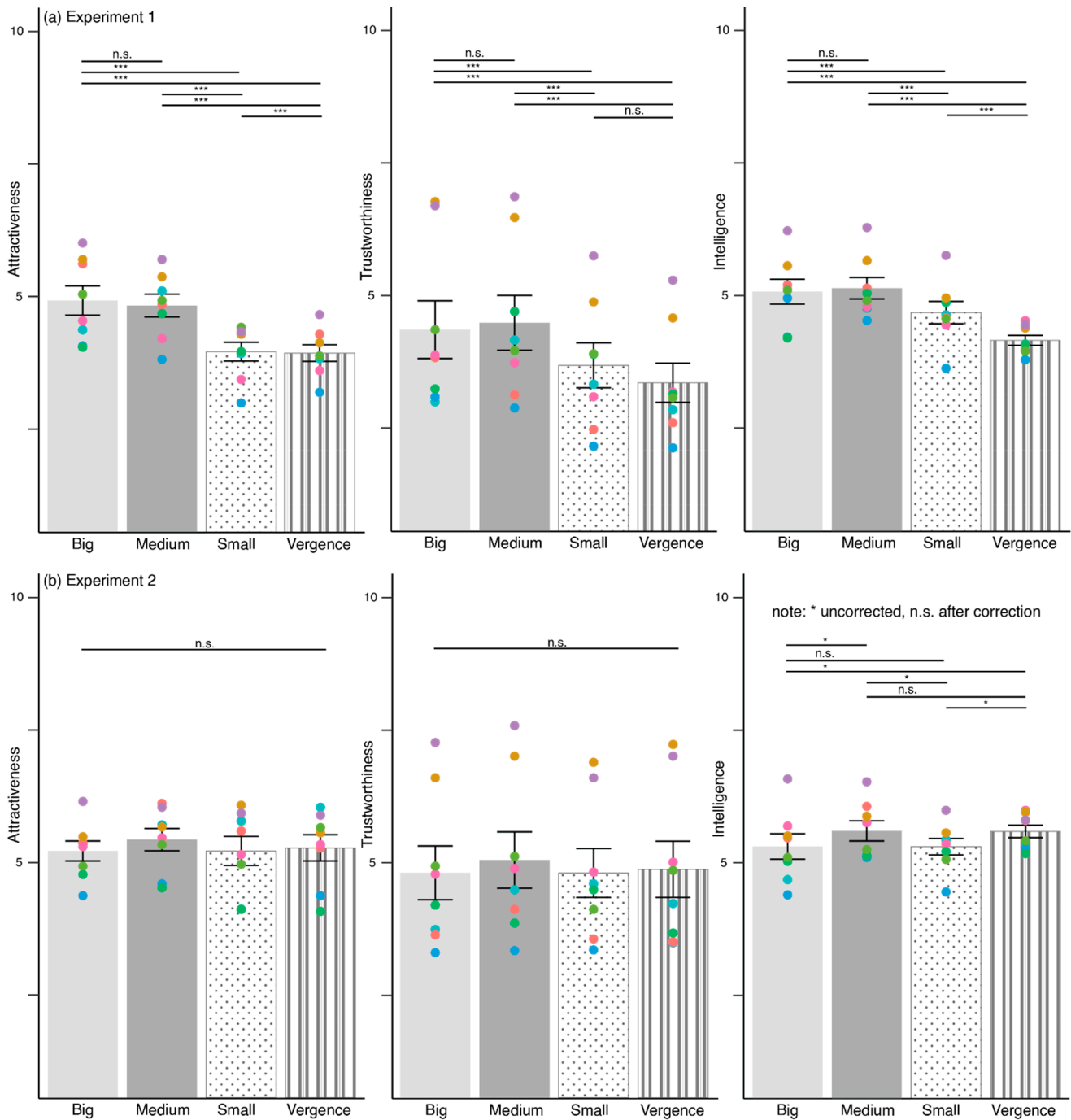


Figure 5. Attractiveness, trustworthiness, and intelligence ratings in (a) Experiment 1, and (b) Experiment 2 across big, medium, small, and vergence pupils. The colored scatter corresponds to each eye model. Error bars depict standard error of means, S.E.M. *** $p < 0.001$, * $p < 0.05$. Intelligence ratings in Experiment 2 are not significant after Bonferroni-correction ($p < 0.008$). n.s.—non-significant.

Table 1. Post-hoc comparisons for attractiveness, trustworthiness, and intelligence ratings in Experiment 1.

Factor	Comparison Pair	<i>t</i> -Stat	<i>df</i>	<i>SD</i>	<i>p</i>	Cohen's <i>d</i>
Attractiveness	Big (<i>M</i> = 4.34, <i>SD</i> = 1.68) & Medium (<i>M</i> = 4.50, <i>SD</i> = 1.51)	−1.09	235	2.20	0.28	0.07
	Big & Small (<i>M</i> = 3.71, <i>SD</i> = 1.56)	4.76	235	2.05	<0.001	0.31
	Big & Vergence (<i>M</i> = 3.33, <i>SD</i> = 1.63)	8.45	235	1.83	<0.001	0.55
	Medium & Small	7.12	235	1.71	<0.001	0.46
	Medium & Vergence	9.09	235	1.97	<0.001	0.59
	Small & Vergence	3.00	235	1.90	0.003	0.20
Trustworthiness	Big (<i>M</i> = 4.93, <i>SD</i> = 1.63) & Medium (<i>M</i> = 4.83, <i>SD</i> = 1.40)	0.80	235	1.82	0.42	0.05
	Big & Small (<i>M</i> = 3.95, <i>SD</i> = 1.67)	7.33	235	2.05	<0.001	0.48
	Big & Vergence (<i>M</i> = 3.93, <i>SD</i> = 1.56)	8.36	235	1.83	<0.001	0.54
	Medium & Small	7.19	235	1.88	<0.001	0.47
	Medium & Vergence	8.15	235	1.70	<0.001	0.53
	Small & Vergence	0.17	235	1.78	0.87	0.01
Intelligence	Big (<i>M</i> = 5.08, <i>SD</i> = 1.40) & Medium (<i>M</i> = 5.16, <i>SD</i> = 1.30)	−0.72	235	1.76	0.47	0.05
	Big & Small (<i>M</i> = 4.69, <i>SD</i> = 1.34)	3.60	235	1.63	<0.001	0.23
	Big & Vergence (<i>M</i> = 4.16, <i>SD</i> = 1.60)	7.79	235	1.81	<0.001	0.51
	Medium & Small	4.26	235	1.68	<0.001	0.28
	Medium & Vergence	8.69	235	1.77	<0.001	0.57
	Small & Vergence	4.65	235	1.77	<0.001	0.30

There was a significant main effect for trustworthiness ratings, $F(3, 705) = 40.87$, $p < 0.001$, $\eta_p^2 = 0.15$. Post-hoc pairwise comparison was conducted with Bonferroni-correction ($p < 0.008$) and presented in Table 1. Both big and medium pupils looked more trustworthy than small pupils and vergence.

The main effect for intelligence ratings was statistically different, $F(3, 705) = 32.57$, $p < 0.001$, $\eta_p^2 = 0.12$. Post-hoc pairwise comparison using Bonferroni-correction ($p < 0.008$) was performed. The comparisons indicated that both big and medium pupils appeared more intelligent than small pupils and vergence. Small pupils also looked more intelligent than vergence.

Figure 5b shows the ratings for attractiveness, trustworthiness, and intelligence for Experiment 2. There were no differences for attractiveness ratings, $F(2.31, 244.63) = 0.65$, $p = 0.58$, $\eta_p^2 = 0.006$, and trustworthiness ratings, $F(3, 705) = 0.91$, $p = 0.44$, $\eta_p^2 = 0.009$.

There was a main effect for intelligence ratings, $F(3, 705) = 3.20$, $p = 0.024$, $\eta_p^2 = 0.03$. Post-hoc pairwise comparison was conducted using Bonferroni-correction ($p < 0.008$) and presented in Table 2. There were statistical differences in intelligence ratings between different pupil sizes. However, none of these comparisons survived Bonferroni-correction.

Table 2. Post-hoc comparisons for intelligence ratings in Experiment 2.

Factor	Comparison Pair	<i>t</i> -Stat	<i>df</i>	<i>SD</i>	<i>p</i>	Cohen's <i>d</i>
Intelligence	Big (<i>M</i> = 5.32, <i>SD</i> = 1.00) & Medium (<i>M</i> = 5.61, <i>SD</i> = 1.17)	−2.12	235	1.39	0.036	0.21
	Big & Small (<i>M</i> = 5.32, <i>SD</i> = 1.03)	0.04	235	1.21	0.97	0.004
	Big & Vergence (<i>M</i> = 5.60, <i>SD</i> = 1.17)	−2.29	235	1.27	0.024	0.22
	Medium & Small	−2.22	235	1.35	0.028	0.22
	Medium & Vergence	0.03	235	1.55	0.98	0.003
	Small & Vergence	−2.29	235	1.29	0.024	0.22

In summary, the pupils significantly influenced the ratings for attractiveness, trustworthiness, and intelligence for Experiment 1. Small and vergence pupils received the

lowest ratings as compared to big and medium pupils. This result was not replicated in Experiment 2. Instead, there was only a significant main effect of perceived intelligence.

3.3. Unrealistic Pupils Influence Perceived Realism and Familiarity

Figure 6a illustrates the ratings for realism and familiarity in Experiment 1. Realism ratings were different, $F(2.83, 665.13) = 99.27, p < 0.001, \eta_p^2 = 0.30$. Post-hoc pairwise comparison using Bonferroni-correction was conducted ($p < 0.008$) and the results depicted in Table 3. The results illustrated that both big and medium pupils appeared more realistic than small pupils and vergence. Small pupils did not differ from vergence since the comparison did not survive Bonferroni-correction.

Table 3. Post-hoc comparisons for realism and familiarity ratings in Experiment 1.

Factor	Comparison Pair	<i>t</i> -Stat	<i>df</i>	<i>SD</i>	<i>p</i>	Cohen's <i>d</i>
Realism	Big ($M = 5.13, SD = 2.17$) & Medium ($M = 5.22, SD = 1.90$)	−0.58	235	2.23	0.56	0.04
	Big & Small ($M = 3.56, SD = 2.11$)	10.14	235	2.37	<0.001	0.66
	Big & Vergence ($M = 3.26, SD = 2.11$)	11.42	235	2.51	<0.001	0.74
	Medium & Small	13.35	235	1.90	<0.001	0.87
	Medium & Vergence	13.67	235	2.19	<0.001	0.89
	Small & Vergence	2.15	235	2.13	0.03	0.14
Familiarity	Big ($M = 3.08, SD = 1.89$) & Medium ($M = 3.12, SD = 1.84$)	−0.30	235	1.87	0.77	0.02
	Big & Small ($M = 2.40, SD = 1.47$)	5.57	235	1.86	<0.001	0.36
	Big & Vergence ($M = 2.24, SD = 1.61$)	7.17	235	1.80	<0.001	0.47
	Medium & Small	6.47	235	1.69	<0.001	0.42
	Medium & Vergence	7.58	235	1.77	<0.001	0.49
	Small & Vergence	1.50	235	1.67	0.13	0.10

The main effect for familiarity ratings was significant, $F(3, 705) = 30.62, p < 0.001, \eta_p^2 = 0.12$. Post-hoc pairwise comparison using Bonferroni-correction ($p < 0.008$) was conducted and the results shown in Table 3. Again, both big and medium pupils looked more familiar than small pupils and vergence.

Figure 6b visualizes the realism and familiarity ratings for Experiment 2. The main effect for realism ratings was not significant, $F(3, 705) = 2.18, p = 0.09, \eta_p^2 = 0.02$. There was no significant main effect for familiarity ratings, $F(3, 705) = 1.32, p = 0.27, \eta_p^2 = 0.01$. In short, big and medium pupils appeared more realistic and familiar than small and vergence pupils for Experiment 1. This effect was absent in Experiment 2.

3.4. Pupils Do Not Influence Perceived Arousal, and Unrealistic Pupils Influence Perceived Valence

Figure 7a illustrates arousal and valence ratings in Experiment 1. There were no differences for arousal ratings, $F(2.69, 631.62) = 1.65, p = 0.18, \eta_p^2 = 0.007$. However, there were differences for valence ratings, $F(2.90, 681.25) = 21.62, p < 0.001, \eta_p^2 = 0.08$. Post-hoc pairwise comparison with Bonferroni-correction ($p < 0.008$) was run, and the results visualized in Table 4. The comparisons revealed that both big and medium pupils received greater valence ratings than small pupils and vergence.

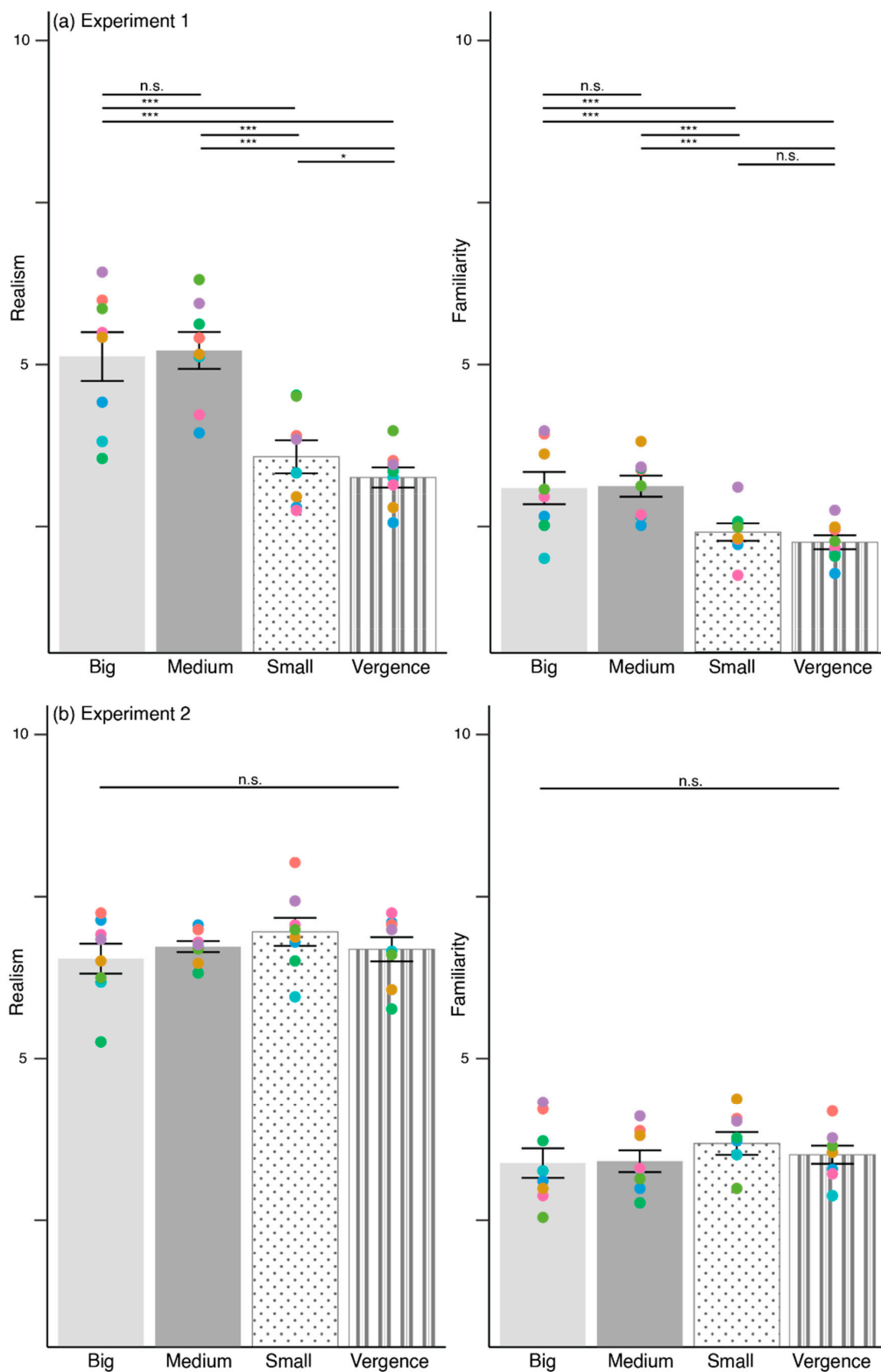


Figure 6. Realism and familiarity ratings in (a) Experiment 1, and (b) Experiment 2 across big, medium, small, and vergence pupils. The colored scatter corresponds to each eye model. Error bars depict standard error of means, S.E.M. *** $p < 0.001$, * $p < 0.05$, n.s.—non-significant.

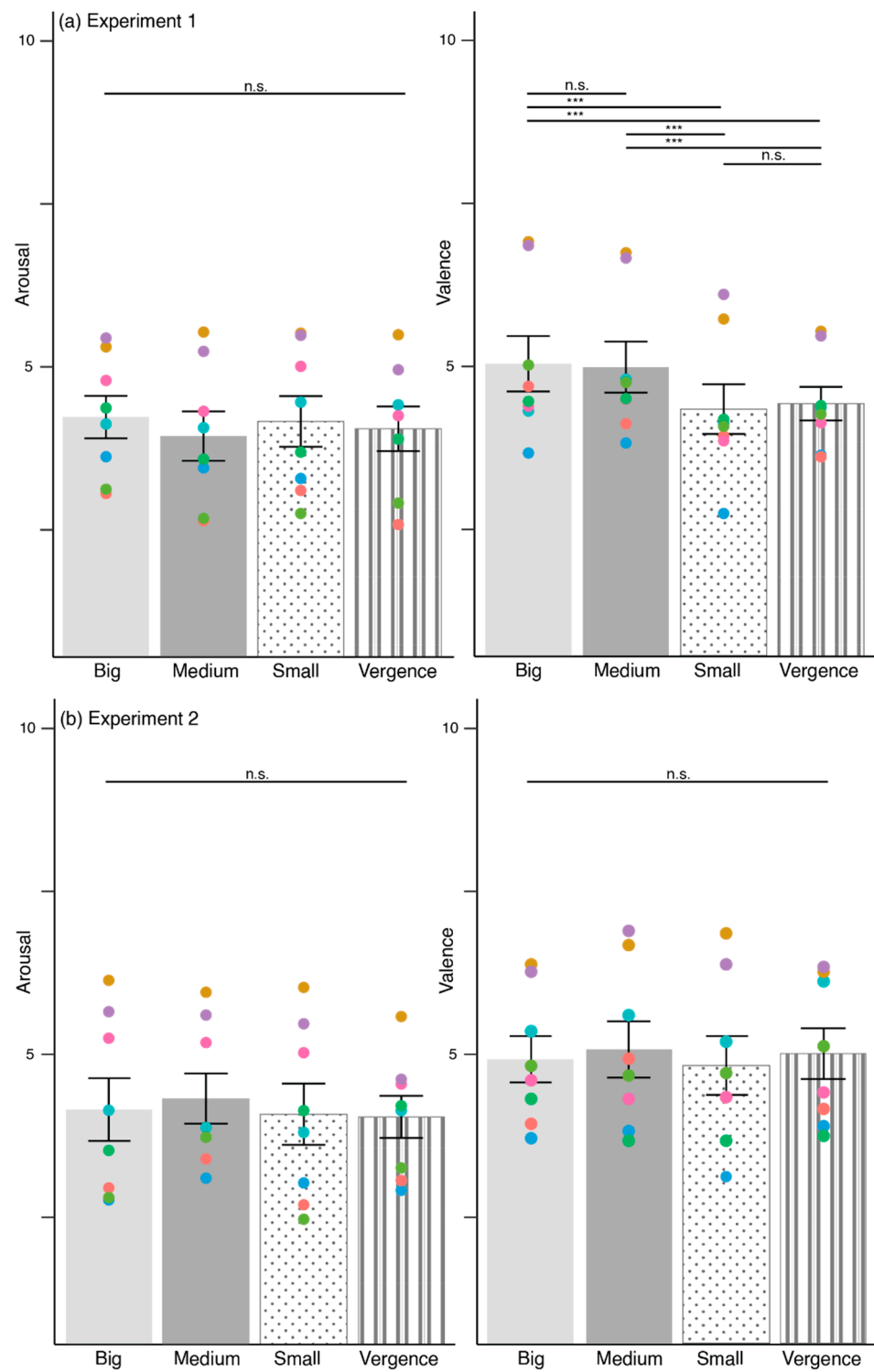


Figure 7. Arousal and valence ratings in (a) Experiment 1, and (b) Experiment 2 across big, medium, small, and vergence pupils. The colored scatter corresponds to each eye model. Error bars depict standard error of means, S.E.M. *** $p < 0.001$, n.s.—non-significant.

Table 4. Post-hoc comparisons for valence ratings in Experiment 1.

Factor	Comparison Pair	<i>t</i> -Stat	<i>df</i>	<i>SD</i>	<i>p</i>	Cohen's <i>d</i>
Valence	Big (<i>M</i> = 5.03, <i>SD</i> = 1.32) & Medium (<i>M</i> = 5.00, <i>SD</i> = 1.25)	0.27	235	1.67	0.79	0.02
	Big & Small (<i>M</i> = 4.35, <i>SD</i> = 1.42)	5.81	235	1.79	<0.001	0.38
	Big & Vergence (<i>M</i> = 4.41, <i>SD</i> = 1.37)	5.55	235	1.70	<0.001	0.36
	Medium & Small	6.03	235	1.65	<0.001	0.39
	Medium & Vergence	4.91	235	1.83	<0.001	0.32
	Small & Vergence	−0.59	235	1.59	0.55	0.04

Figure 7b pictures the arousal and valence ratings for Experiment 2. There was no significant main effect of arousal ratings, $F(2.70, 286.03) = 0.78$, $p = 0.50$, $\eta_p^2 = 0.007$. The main effect for valence ratings was not significant, $F(2.74, 290.48) = 0.38$, $p = 0.38$, $\eta_p^2 = 0.009$.

In conclusion, pupils did not influence arousal ratings in both experiments. However, the pupils influenced valence ratings in Experiment 1. Specifically, small and vergence pupils appeared more negatively in valence than big and medium pupils. This was not replicated in Experiment 2.

4. Discussion

Some studies show that a person's characteristics, such as attractiveness or trustworthiness, could be perceived by looking at the eyes and the pupils [16,17]. In our study using only eyes with neutral expressions, we found large rating differences for exaggerated pupils (Experiment 1), but these differences were mostly absent for realistically looking pupils (Experiment 2). Pupils did not influence perceived sex and age, and they also did not systematically affect ratings of attractiveness, trustworthiness, intelligence, realism, familiarity, and valence. Contrary to the literature, we found no evidence of specific directional associations between pupils and ratings (i.e., big pupils denote higher ratings). Also, participants in both experiments could not perceive a person's arousal from looking at the pupils. Therefore, it seems that the pupils, although when they are exaggerated could sometimes influence the perceived characteristics of another person, provide information mainly when interpreted in relation to other characteristics of facial expression.

4.1. How the Pupils Influence Perceived Characteristics of a Person

It seems unlikely that the pupils can influence the perceived sex and age of a person. In fact, we found no differences in sex and age ratings between the pupils. From the literature, biological sex influences eye size [6,37]. However, biological sex is unrelated to pupil size [38,39]. Age is typically perceived via features such as wrinkles [9,40]. However, pupil size is related to age. Older people have smaller pupils than younger people [41]. This decrease in pupil size is also very small [42] and difficult to perceive. We speculate that one could perceive age if the pupils appear cloudy (i.e., cataracts). This is due to the fact that cataracts are common in the older population [43]. Nonetheless, it is unlikely that sex and age can be perceived from changing pupil sizes.

Other characteristics about a person may be perceived by looking at the pupils, which could be driven by more complex processes. Our data show that the perceived attractiveness, trustworthiness, intelligence, realism, familiarity, and valence, can be influenced by pupil size when looking at neutral eyes. However, our results do not show that pupil sizes alone influence how a person reads various characteristics about another person by looking at the eyes.

Kret and De Dreu [17] showed that big (i.e., dilating) pupils appeared more attractive and trustworthy than small (i.e., constricting) pupils. In their experiment, big and small pupils differed by over 80% pupil area. If Kret and De Dreu's results could be explained by pupil size differences, then we should also observe similar results between big and medium pupils in Experiment 1, since big and medium pupils differed by 100% in area. We found

no differences between these two exaggerated pupils (Experiment 1). We also did not find any differences between big and small realistic pupils (Experiment 2) and they differed by twice the size (+100% area). Thus, it is unlikely that size alone can explain why the ratings differed from our study.

A key difference between our results and those from Kret and De Dreu's study is motion. Although both studies used neutral eyes, our stimuli were static images, as compared to dynamic pupil changes. This hints that the changing of pupil sizes influences our perception regarding attractiveness and trustworthiness of another person. As a final remark, our results for exaggerated pupils changed in the same direction as those in Kret et al. [17], despite being static images. Hence, it is possible that there may be an interaction between static pupil sizes and motion from dynamically changing pupils, which contribute to how we read different characteristics about a person by looking at their eyes.

4.2. Ecological Validity of Image Manipulation in Experiments

The directional associations between pupils and ratings shown in previous studies [12,17,44] were missing from our study. Instead, in Experiment 1, eyes with small and vergence pupils were rated much lower than big and medium pupils in attractiveness, trustworthiness, intelligence, realism, familiarity, and valence. This finding was not replicated in Experiment 2 with more realistic pupils. This finding raises the question concerning to which extent ecological validity in stimulus preparation influences the generalizability of the results.

One explanation of the differences in our results might be related to the *uncanny valley* effect. This effect refers to evoked feelings of uneasiness when an artificial agent or imagery looks almost human-like but not perfectly [45]. It might be that in Experiment 1, the small pupils and the simulated vergence evoke *uncanny* vibes, lessening the trustworthiness of the eyes in comparison to big pupils. Importantly, big pupils also looked unrealistic, but was not susceptible to the uncanny valley effect. Participants could have rated big pupils as less unrealistic, but not considered them uncanny based on a general understanding of pupillary response. Big pupils are uncommon in well-lit conditions for most every-day situations. The pupils constrict and become comparably small in response to light or changes in illumination [46,47]. Therefore, the exaggerated big pupils, although appearing unrealistic in Experiment 1, were not prone to the uncanny valley effect.

Laboratory experiments often take place in highly controlled environments so that manipulations can be accounted for. However, lab environments hardly mimic the real-world. When interacting in the real world, we often integrate multisensory cues to minimize communication errors [48,49]. These multisensory cues are often missing in experiments to simplify the number of variables needed to be manipulated. As a result, realistic stimuli often produce smaller effects compared to artificial/schematic ones [50]. In the context of our study, the eye-stimuli are also manipulated in unrealistic ways in existing literature, either by changing pupil sizes across an arbitrary duration [16,17], or by artificially modifying the sizes [12]. However, these manipulations are not replicable in the real world.

Pupillary changes are dynamic and unpredictable in the real world. The pupils can change very quickly in response to minor fluctuations in luminance, cognitive processes, and arousal [42,51,52]. Some studies suggest that our eyes can sense changes in another person's pupil sizes through mimicking the pupil size of another person or animal during social interactions [16,53,54]. However, these studies cannot reliably isolate the influence of luminance on pupil dynamics [55,56]. In fact, pupil dynamics may not be easily seen in the real world since most (79%) people have brown irises compared to blue (10%) and other colors across the world [57]. There is also a spectrum of different shades for brown irises. Thus, it is difficult to say whether sensation of another person's pupil size is possible despite the ever-changing luminance in the environment since there is a high chance that most people do not even see another person's pupils due to the iris color.

It is unlikely that we can process changes in another person's pupils unconsciously or consciously report the pupillary changes in real-world interactions. This is due to the

fact that we do not make constant eye contact with whom we interact with. Eye contact can vary depending on the physical distance between people and the interaction's phase. For example, eye contact is shorter and less frequent between the opposite sex, or when members are very close together [58]. In contrast, eye contact is more common when someone is listening to a speaker [59]. It is also unlikely that we pay close attention and look intently into another person's eyes throughout the interaction as doing so causes the other person to feel uneasy. Since we do not make constant eye contact with others, it is less likely that we unconsciously process another person's pupillary changes or become aware of such changes during real-world interactions. Hence, we can speculate that the impact of our measures may not replicate to real-world social interactions, since participants in our experiments were aware of the differently sized pupils and vergence when rating characteristics.

Lastly, an important aspect to consider is whether the stimulus preparation could be meaningful in drawing parallel conclusions about the real world. To that, some researchers have called for a clearer context-driven approach in defining ecological validity [60], where research questions should be defined as context-specific questions. In any case, future experiments should adopt realistic and multisensory stimuli, whilst formulating specific contexts, to further investigate the pupils' role in influencing how we perceive characteristics about another person by looking at their eyes.

4.3. *Perceiving Arousal from the Pupils*

Pupil dilation indicates arousal when pupil size changes are tracked in an observer [61]. The pupils also dilate when we are emotionally stimulated [62]. It seems that we may not be good at perceiving a person's emotional arousal when looking at their pupils. This is since we found no differences in arousal ratings for both experiments. One criticism to our claim could be since the exaggerated pupils appeared absurd in Experiment 1. However, we did not find any differences when using more realistic pupils in Experiment 2. The findings may change if we had presented emotionally expressive eyes. However, this would suggest that the perceived emotional arousal was extracted by integrating information about the eye expression and the pupil size. In short, it is likely that we are bad at perceiving a person's emotional arousal from looking at their pupils.

Additional visual signals besides the pupils are required to perceive a person's arousal, specifically the sexual arousal. The pupils dilate when a person is sexually aroused. The pupils dilate in response to sexual attraction [63]. Consequently, a person with larger pupils is perceived as more attractive [44,64]. One may even evaluate a stranger's pupils in online dating to pick their potential partners [65]. These studies show that pupil sizes influence perceived arousal (in terms of sexual arousal) when a full face is visible. This differs from our study where only the eyes were shown. Since we did not measure perceived sexual arousal, it remains possible that sexual arousal could be perceived by looking at the pupils in the eyes.

4.4. *The Pupils as Important Cues in Interactions*

It is important to investigate how the pupils influence what we perceive about another person. The pupils tell us where the eyes are gazing [66,67], so they help us understand how the eyes inform mutual interaction [68]. In such interactions, pupil dilations communicate trust [69], eye blinks signal the transition of conversational topics [70], gaze patterns can indicate specific cognitive processes that are relevant for learning [71] and recognizing eye movement patterns facilitates the recognition of facial expressions [72]. Teachers who gaze into the students' eyes during teaching are perceived to show greater interests in the student's learning [73]. Gazing at students also helps capture the students' attention [74] and fosters an interpersonal connection to their students, which is critical for the quality of teaching and learning [75,76]. With the recent surge in online-schooling, the frequency of making eye contact has fallen dramatically, creating severe problems in teaching as teachers must be able to observe their students' attention to create an effective and efficient learning

environment. In on-site teaching contexts, teachers can instantly pool information from many pupils to infer the direction of a group's collective gaze [77]. In online learning, this is not possible (since the webcam is not embedded in the screen), so alternative measures need to be considered in online education, such as mapping the teachers gaze point on the material for guiding the students [78] or visualizing the aggregated gaze points of students [79].

4.5. Limitations and Future Directions

The current study investigates the pupils in neutral static eye images. Thus, the results are specific only to static eyes without expressions. It is known that we rely on the eyes to identify fear and surprise [80,81]. Dynamic changes in the eye region also influence whether one correctly recognizes certain emotional expressions [82]. Since the eyes also convey different emotions, it would be interesting to investigate how the interpretation of the current study changes, when eye emotions are introduced. Future studies could investigate the influence of varying pupils of either different static emotional eye expressions or dynamic eye expressions on the perceived characteristics of another person by looking at the eyes. The current sample sizes differ between the two experiments. Future experiments should aim for more balanced samples, especially when comparing data across experiments.

The addition of the first Purkinje image makes the eyes appear more lifelike. This is a technique commonly used in art. One criticism of Experiment 2 was that the reflections could have reduced the overall perceivable pupil size. According to the literature, the presence and absence of these reflections do not alter the perceived gaze of a painted portrait. It only affects the realism [83]. Therefore, subsequent studies can compare whether the first Purkinje images reduce the overall perceived pupil size.

Cultural factors influence how often a person looks at the eyes. Eastern cultures have smaller emphasis on the eyes than the western culture [84,85]. We did not query participants' ethnicity. Participant recruitment was also of convenient means (i.e., social media, friends, families). We also did not test eyes from other ethnicities. Hence, our conclusions are limited to the western culture.

It is without a doubt challenging in capturing the pupil dynamics in a controlled and reliable manner so that they mimic real-world behaviors. Nevertheless, future studies should strive for more ecological valid stimuli, such as recording real pupil dynamics, and presenting them as stimuli. Studies should also define context-specific problems to solve, so that ecological validity could be captured more precisely. Alternatively, the study could be conducted on real-person interactions where the experimental conditions elicit dilated or constricted pupils.

5. Conclusions

The eyes are important in communication since they convey information about our cognitive and affective states. We investigated the effects of pupil size and vergence on how an observer reads various characteristics of another person by looking at the person's neutral expression eyes. In exaggerated pupils, we found that small pupils and vergence were rated lower in attractiveness, trustworthiness, intelligence, realism, familiarity, and valence than large and medium pupils. These findings were absent for realistic pupils. The exaggeration could have induced some form of uncanniness, causing participants to rate them lower than realistic eyes. Regardless, our results indicate that the pupils may influence how we perceive a person's characteristics, even when we look at emotionally neutral eyes. Therefore, the findings suggest that pupil size and vergence may be regarded as important social signals in subconscious social interaction processes. Importantly, future experiments need to pay stronger emphasis on ecologically valid stimuli when investigating the pupils and its relationship regarding the perceived characteristics of a person.

Author Contributions: Conceptualization, A.H. and W.K.L.; methodology, W.K.L.; validation, W.K.L. and M.S.; formal analysis, W.K.L.; investigation, W.K.L.; resources, W.K.L., M.S. and A.H.; data curation, W.K.L.; writing—original draft preparation, W.K.L.; writing—review and editing, A.H.,

W.K.L. and M.S.; visualization, W.K.L.; project administration, W.K.L.; funding acquisition, A.H. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Deutsche Forschungsgemeinschaft Projektnummer 425867974 (DFG, German Research Foundation) and is part of the “SPP 2199: Skalierbare Interaktionsparadigmen für allgegenwärtige Rechnerumgebungen“ (Priority Program SPP2199 Scalable Interaction Paradigms for Pervasive Computing Environments).

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 data and analyses scripts associate with this manuscript are available from: Lau, Wee Kiat (2022): The role of realism in perceiving neutral expression eyes. Open Access Repository der Universität Ulm und Technischen Hochschule Ulm. <http://dx.doi.org/10.18725/OPARU-42603> (accessed on 22 June 2022).

Acknowledgments: The authors would like to thank the bachelor students, K.M. and R.C., and the students from the psychology courses for participation and for advertising the experiment to their friends and families.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

We tested whether the statistical assumptions were violated prior to our analyses. The results of these tests were reported below.

Appendix A.1. Analysis 3.1, Experiment 1: Sex and Age

The sphericity assumption for sex ratings was violated, Mauchly's $W = 0.91$, $\chi^2 = 23.45$, $p < 0.001$, and the subsequent analysis adopted the Greenhouse-Geisser correction. The sphericity assumption for age ratings was not violated, Mauchly's $W = 0.97$, $\chi^2 = 6.87$, $p = 0.23$. No correction was applied for the subsequent analysis.

Appendix A.2. Analysis 3.1, Experiment 2: Sex and Age

The sphericity assumption for sex ratings was violated, Mauchly's $W = 0.87$, $\chi^2 = 14.48$, $p = 0.01$. The Greenhouse-Geisser correction was applied. The sphericity assumption for age rating was not violated, Mauchly's $W = 0.98$, $\chi^2 = 2.40$, $p = 0.79$. No additional correction method was used.

Appendix A.3. Analysis 3.2, Experiment 1: Attractiveness, Trustworthiness, and Intelligence

The sphericity assumption for attractiveness ratings was violated, Mauchly's $W = 0.89$, $\chi^2 = 27.29$, $p < 0.001$. The Greenhouse-Geisser correction was applied. The sphericity assumption for trustworthiness ratings was not violated, Mauchly's $W = 0.96$, $\chi^2 = 10.44$, $p = 0.06$. No further correction methods were applied. The sphericity assumption for intelligence ratings was not violated, Mauchly's $W = 0.98$, $\chi^2 = 4.28$, $p = 0.51$. No further correction was applied.

Appendix A.4. Analysis 3.2, Experiment 2: Attractiveness, Trustworthiness, and Intelligence

The sphericity assumption for attractiveness ratings was violated, Mauchly's $W = 0.66$, $\chi^2 = 43.12$, $p < 0.001$. The Greenhouse-Geisser correction was applied. The sphericity assumption for trustworthiness ratings was not violated, Mauchly's $W = 0.95$, $\chi^2 = 5.37$, $p = 0.37$. No additional correction was required. The sphericity assumption for intelligence ratings was not violated, Mauchly's $W = 0.92$, $\chi^2 = 9.22$, $p = 0.10$. No correction was required.

Appendix A.5. Analysis 3.3, Experiment 1: Realism and Familiarity

The sphericity assumption for realism ratings was violated, Mauchly's $W = 0.91$, $\chi^2 = 21.61$, $p < 0.001$. The Greenhouse-Geisser correction was applied. The sphericity assumption for familiarity ratings was not violated, Mauchly's $W = 0.97$, $\chi^2 = 6.10$, $p = 0.30$. No additional correction was used.

Appendix A.6. Analysis 3.3, Experiment 2: Realism and Familiarity

The sphericity assumption for realism was not violated, Mauchly's $W = 0.98$, $\chi^2 = 2.17$, $p = 0.83$. No further corrections were applied. The sphericity assumption for familiarity ratings was not violated, Mauchly's $W = 0.94$, $\chi^2 = 6.52$, $p = 0.28$. No additional corrections were applied.

Appendix A.7. Analysis 3.4, Experiment 1: Arousal and Valence

The sphericity assumption for arousal ratings was violated, Mauchly's $W = 0.82$, $\chi^2 = 44.07$, $p < 0.001$. The Greenhouse-Geisser correction was applied. The sphericity assumption for valence ratings was violated, Mauchly's $W = 0.94$, $\chi^2 = 13.42$, $p = 0.02$. The Greenhouse-Geisser correction was used.

Appendix A.8. Analysis 3.4, Experiment 2: Arousal and Valence

The sphericity assumption for arousal ratings was violated, Mauchly's $W = 0.86$, $\chi^2 = 16.00$, $p = 0.007$. The Greenhouse-Geisser correction was applied for the next analysis. The sphericity assumption for valence ratings was violated, Mauchly's $W = 0.87$, $\chi^2 = 15.17$, $p = 0.01$. The Greenhouse-Geisser correction was used.

Appendix B.

We ran additional correlational analyses between attractiveness, trustworthiness, and intelligence ratings for each experiment. There were no consistencies in the correlations across the three variables.

Table A1. Pearson correlations for Experiment 1.

Factor		Attractiveness				Trustworthiness				Intelligence			
		B	M	S	V	B	M	S	V	B	M	S	V
Attractiveness	B	1											
	M	0.05	1										
	S	0.20	0.38	1									
	V	0.39	0.22	0.29	1								
Trustworthiness	B	0.63	0.13	0.20	0.31	1							
	M	0.20	0.45	0.18	0.28	0.28	1						
	S	0.18	0.21	0.56	0.35	0.23	0.26	1					
	V	0.19	0.26	0.32	0.57	0.34	0.35	0.40	1				
Intelligence	B	0.61	-0.02	0.08	0.27	0.58	0.22	0.11	0.19	1			
	M	0.02	0.40	0.09	0.06	0.18	0.42	0.07	0.19	0.15	1		
	S	0.25	0.13	0.48	0.28	0.28	0.22	0.44	0.33	0.29	0.20	1	
	V	0.21	0.18	0.24	0.63	0.25	0.25	0.31	0.56	0.28	0.27	0.28	1

B—big, M—medium, S—small, V—vergence. Blue: $p < 0.01$, Yellow: $p < 0.05$.

Table A2. Pearson correlations for Experiment 2.

Factor		Attractiveness				Trustworthiness				Intelligence			
		B	M	S	V	B	M	S	V	B	M	S	V
Attractiveness	B	1											
	M	0.06	1										
	S	0.07	0.40	1									
	V	0.52	−0.05	−0.11	1								
Trustworthiness	B	0.53	0.14	0.18	0.26	1							
	M	0.17	0.34	0.15	0.09	0.28	1						
	S	0.05	0.12	0.43	0.01	0.12	0.23	1					
	V	0.25	0.19	0.13	0.53	0.39	0.19	0.13	1				
Intelligence	B	0.64	0.04	0.13	0.27	0.46	0.10	0.05	0.22	1			
	M	0.20	0.58	0.29	−0.12	0.08	0.44	0.16	−0.02	0.19	1		
	S	0.27	−0.01	0.38	0.08	0.12	0.12	0.30	0.13	0.30	0.25	1	
	V	0.21	0.14	0.05	0.48	0.18	0.14	0.10	0.41	0.32	0.11	0.32	1

B—big, M—medium, S—small, V—vergence. Blue: $p < 0.01$, Yellow: $p < 0.05$.

References

- Batki, A.; Baron-Cohen, S.; Wheelwright, S.; Connellan, J.; Ahluwalia, J. Is there an innate gaze module? Evidence from human neonates. *Infant Behav. Dev.* **2000**, *23*, 223–229. [\[CrossRef\]](#)
- Taylor, M.J.; Edmonds, G.E.; McCarthy, G.; Allison, T. Eyes first! Eye processing develops before face processing in children. *Neuroreport* **2001**, *12*, 1671–1676. [\[CrossRef\]](#) [\[PubMed\]](#)
- Todorov, A.; Duchaine, B. Reading trustworthiness in faces without recognizing faces. *Cogn. Neuropsychol.* **2008**, *25*, 395–410. [\[CrossRef\]](#)
- Willis, J.; Todorov, A. First impressions: Making up your mind after a 100-ms exposure to a face. *Psychol. Sci.* **2006**, *17*, 592–598. [\[CrossRef\]](#)
- Todorov, A.; Pakrashi, M.; Oosterhof, N.N. Evaluating faces on trustworthiness after minimal time exposure. *Soc. Cogn.* **2009**, *27*, 813–833. [\[CrossRef\]](#)
- Midelfart, A. Women and men—same eyes? *Acta Ophthalmol. Scand.* **1996**, *74*, 589–592. [\[CrossRef\]](#)
- Tipples, J.; Atkinson, A.P.; Young, A.W. The eyebrow frown: A salient social signal. *Emotion* **2002**, *2*, 288. [\[CrossRef\]](#)
- Gunnery, S.D.; Ruben, M.A. Perceptions of Duchenne and non-Duchenne smiles: A meta-analysis. *Cogn. Emot.* **2016**, *30*, 501–515. [\[CrossRef\]](#)
- Ganel, T.; Goodale, M.A. The effect of smiling on the perceived age of male and female faces across the lifespan. *Sci. Rep.* **2021**, *11*, 23020. [\[CrossRef\]](#)
- Flament, F.; Pierre, J.; Delhommeau, K.; Adam, A. How a working day-induced-tiredness may alter some facial signs in differently-aged Caucasian women. *Int. J. Cosmet. Sci.* **2017**, *39*, 467–475. [\[CrossRef\]](#)
- Baron-Cohen, S.; Wheelwright, S.; Jolliffe, A.T. Is there a “language of the eyes”? Evidence from normal adults, and adults with autism or Asperger syndrome. *Vis. Cogn.* **1997**, *4*, 311–331. [\[CrossRef\]](#)
- Harrison, N.A.; Wilson, C.E.; Critchley, H.D. Processing of observed pupil size modulates perception of sadness and predicts empathy. *Emotion* **2007**, *7*, 724. [\[CrossRef\]](#) [\[PubMed\]](#)
- Milcent, A.-S.; Geslin, E.; Kadri, A.; Richir, S. Expressive Virtual Human: Impact of expressive wrinkles and pupillary size on emotion recognition. In Proceedings of the 19th ACM International Conference on Intelligent Virtual Agents, Paris, France, 2–5 July 2019; pp. 215–217.
- Cunningham, M.R.; Barbee, A.P.; Pike, C.L. What do women want? Facialmetric assessment of multiple motives in the perception of male facial physical attractiveness. *J. Person. Soc. Psychol.* **1990**, *59*, 61. [\[CrossRef\]](#)
- Przylipiak, M.; Przylipiak, J.; Terlikowski, R.; Lubowicka, E.; Chrostek, L.; Przylipiak, A. Impact of face proportions on face attractiveness. *J. Cosmet. Dermatol.* **2018**, *17*, 954–959. [\[CrossRef\]](#)
- Kret, M.; Fischer, A.; De Dreu, C.K. Pupil mimicry correlates with trust in in-group partners with dilating pupils. *Psychol. Sci.* **2015**, *26*, 1401–1410. [\[CrossRef\]](#)
- Kret, M.; De Dreu, C.K. The power of pupil size in establishing trust and reciprocity. *J. Exp. Psychol. Gen.* **2019**, *148*, 1299. [\[CrossRef\]](#)
- Enright, J. Perspective vergence: Oculomotor responses to line drawings. *Vis. Res.* **1987**, *27*, 1513–1526. [\[CrossRef\]](#)
- McLin, L.N., Jr.; Schor, C.M.; Kruger, P.B. Changing size (looming) as a stimulus to accommodation and vergence. *Vis. Res.* **1988**, *28*, 883–898. [\[CrossRef\]](#)
- Horstmann, G.; Loth, S. The Mona Lisa Illusion—Scientists See Her Looking at Them Though She Isn't. *i-Perception* **2019**, *10*, 2041669518821702. [\[CrossRef\]](#)
- Tolchin, J.G.; Lederman, M.E. Congenital (Infantile) Esotropia: Psychiatric Aspects. *J. Pediatr. Ophthalmol. Strabismus* **1978**, *15*, 160–163. [\[CrossRef\]](#)

22. Hillstrom, A.P.; Yantis, S. Visual motion and attentional capture. *Percept. Psychophys.* **1994**, *55*, 399–411. [[CrossRef](#)] [[PubMed](#)]
23. Bradley, M.M.; Lang, P.J. Measuring emotion: The self-assessment manikin and the semantic differential. *J. Behav. Ther. Exp. Psychiatry* **1994**, *25*, 49–59. [[CrossRef](#)]
24. Tsukahara, J.S.; Harrison, T.L.; Engle, R.W. The relationship between baseline pupil size and intelligence. *Cogn. Psychol.* **2016**, *91*, 109–123. [[CrossRef](#)] [[PubMed](#)]
25. Ebner, N.C.; Riediger, M.; Lindenberger, U. FACES—A database of facial expressions in young, middle-aged, and older women and men: Development and validation. *Behav. Res. Methods* **2010**, *42*, 351–362. [[CrossRef](#)] [[PubMed](#)]
26. Lau, W.K. Face Masks Bolsters the Characteristics From Looking at a Face Even When Facial Expressions Are Impaired. *Front. Psychol.* **2021**, *12*, 704916. [[CrossRef](#)] [[PubMed](#)]
27. Van Run, L.; Van den Berg, A. Binocular eye orientation during fixations: Listing’s law extended to include eye vergence. *Vis. Res.* **1993**, *33*, 691–708. [[CrossRef](#)]
28. Mok, D.; Ro, A.; Cadera, W.; Crawford, J.; Vilis, T. Rotation of Listing’s plane during vergence. *Vis. Res.* **1992**, *32*, 2055–2064. [[CrossRef](#)]
29. GmbH, Q. EFS Survey. In *Version Fall 2020*; Questback GmbH: Colgone, Germany, 2021.
30. Mehrabian, A.; Russell, J.A. *An Approach to Environmental Psychology*; MIT Press: Cambridge, MA, USA, 1974.
31. Lau, W.K.; Huckauf, A. Effects of face masks on the appearance of emotional expressions and invariant characteristics. *Open Psychol.* **2021**, *3*, 87–102. [[CrossRef](#)]
32. Abdi, H. Bonferroni Test. In *Encyclopedia of Measurement and Statistics*; Salkind, N.J., Ed.; Sage Publications, Inc.: Thousand Oaks, CA, USA, 2007; pp. 103–107. [[CrossRef](#)]
33. Mauchly, J.W. Significance test for sphericity of a normal n-variate distribution. *Ann. Math. Stat.* **1940**, *11*, 204–209. [[CrossRef](#)]
34. Greenhouse, S.W.; Geisser, S. On methods in the analysis of profile data. *Psychometrika* **1959**, *24*, 95–112. [[CrossRef](#)]
35. Quadflieg, S.; Vermeulen, N.; Rossion, B. Differential Reliance on the Duchenne Marker During Smile Evaluations and Person Judgments. *J. Nonverb. Behav.* **2013**, *37*, 69–77. [[CrossRef](#)]
36. Talamas, S.N.; Mavor, K.I.; Axelsson, J.; Sundelin, T.; Perrett, D.I. Eyelid-openness and mouth curvature influence perceived intelligence beyond attractiveness. *J. Exp. Psychol. Gen.* **2016**, *145*, 603–620. [[CrossRef](#)] [[PubMed](#)]
37. Lim, K.J.; Hyung, S.M.; Youn, D.H. Ocular dimensions with aging in normal eyes. *Korean J. Ophthalmol.* **1992**, *6*, 19–31. [[CrossRef](#)] [[PubMed](#)]
38. Bradley, J.C.; Bentley, K.C.; Mughal, A.I.; Bodhireddy, H.; Young, R.S.; Brown, S.M. The effect of gender and iris color on the dark-adapted pupil diameter. *J. Ocular Pharmacol. Ther.* **2010**, *26*, 335–340. [[CrossRef](#)] [[PubMed](#)]
39. Netto, M.V.; Ambrósio, R.; Wilson, S.E. Pupil Size in Refractive Surgery Candidates. *J. Refract. Surg.* **2004**, *20*, 337–342. [[CrossRef](#)]
40. Nkengne, A.; Bertin, C.; Stamatas, G.; Giron, A.; Rossi, A.; Issachar, N.; Fertil, B. Influence of facial skin attributes on the perceived age of Caucasian women. *J. Eur. Acad. Dermatol. Venereol.* **2008**, *22*, 982–991. [[CrossRef](#)]
41. Birren, J.E.; Casperson, R.C.; Botwinick, J. Age changes in pupil size. *J. Gerontol.* **1950**, *5*, 216–221. [[CrossRef](#)]
42. Winn, B.; Whitaker, D.; Elliott, D.B.; Phillips, N.J. Factors affecting light-adapted pupil size in normal human subjects. *Investig. Ophthalmol. Vis. Sci.* **1994**, *35*, 1132–1137.
43. Thompson, J.; Lakhani, N. Cataracts. *Prim. Care Clin. Off. Pract.* **2015**, *42*, 409–423. [[CrossRef](#)]
44. Hess, E.H.; Seltzer, A.L.; Shlien, J.M. Pupil response of hetero- and homosexual males to pictures of men and women: A pilot study. *J. Abnorm. Psychol.* **1965**, *70*, 165. [[CrossRef](#)]
45. Mathur, M.B.; Reichling, D.B. Navigating a social world with robot partners: A quantitative cartography of the Uncanny Valley. *Cognition* **2016**, *146*, 22–32. [[CrossRef](#)]
46. Ellis, C.J. The pupillary light reflex in normal subjects. *Br. J. Ophthalmol.* **1981**, *65*, 754–759. [[CrossRef](#)] [[PubMed](#)]
47. Campbell, F.W.; Gregory, A.H. Effect of Size of Pupil on Visual Acuity. *Nature* **1960**, *187*, 1121–1123. [[CrossRef](#)] [[PubMed](#)]
48. De Gelder, B. Towards the neurobiology of emotional body language. *Nat. Rev. Neurosci.* **2006**, *7*, 242–249. [[CrossRef](#)] [[PubMed](#)]
49. De Gelder, B.; Bertelson, P. Multisensory integration, perception and ecological validity. *Trends Cogn. Sci.* **2003**, *7*, 460–467. [[CrossRef](#)] [[PubMed](#)]
50. Risko, E.F.; Laidlaw, K.E.; Freeth, M.; Foulsham, T.; Kingstone, A. Social attention with real versus reel stimuli: Toward an empirical approach to concerns about ecological validity. *Front. Hum. Neurosci.* **2012**, *6*, 143. [[CrossRef](#)]
51. Mathôt, S. Pupillometry: Psychology, physiology, and function. *J. Cogn.* **2018**, *1*, 16. [[CrossRef](#)] [[PubMed](#)]
52. Matthews, G.; Middleton, W.; Gilmartin, B.; Bullimore, M.A. Pupillary diameter and cognitive load. *J. Psychophysiol.* **1991**, *5*, 265–271.
53. Axelsson, E.L.; Fawcett, C. Humans’ pupillary contagion extends to cats and dogs. *Soc. Cogn. Affect. Neurosci.* **2021**, *16*, 153–166. [[CrossRef](#)] [[PubMed](#)]
54. Fawcett, C.; Wesevich, V.; Gredebäck, G. Pupillary Contagion in Infancy: Evidence for Spontaneous Transfer of Arousal. *Psychol. Sci.* **2016**, *27*, 997–1003. [[CrossRef](#)] [[PubMed](#)]
55. Derksen, M.; van Alphen, J.; Schaap, S.; Mathot, S.; Naber, M. Pupil Mimicry is the Result of Brightness Perception of the Iris and Pupil. *J. Cogn.* **2018**, *1*, 32. [[CrossRef](#)] [[PubMed](#)]
56. Mathôt, S.; Naber, M. There is no evidence that pupil mimicry is a social phenomenon. *Proc. Natl. Acad. Sci. USA* **2018**, *115*, E11565. [[CrossRef](#)]

57. Dorgaleh, S.; Naghipoor, K.; Barahouie, A.; Dastaviz, F.; Oladnabi, M. Molecular and biochemical mechanisms of human iris color: A comprehensive review. *J. Cell. Physiol.* **2020**, *235*, 8972–8982. [[CrossRef](#)] [[PubMed](#)]
58. Argyle, M.; Dean, J. Eye-Contact, Distance and Affiliation. *Sociometry* **1965**, *28*, 289–304. [[CrossRef](#)]
59. Kleinke, C.L. Gaze and eye contact: A research review. *Psychol. Bull.* **1986**, *100*, 78–100. [[CrossRef](#)]
60. Holleman, G.A.; Hooge, I.T.C.; Kemner, C.; Hessels, R.S. The ‘Real-World Approach’ and Its Problems: A Critique of the Term Ecological Validity. *Front. Psychol.* **2020**, *11*, 721. [[CrossRef](#)] [[PubMed](#)]
61. Wang, C.-A.; Baird, T.; Huang, J.; Coutinho, J.D.; Brien, D.C.; Munoz, D.P. Arousal effects on pupil size, heart rate, and skin conductance in an emotional face task. *Front. Neurol.* **2018**, *9*, 1029. [[CrossRef](#)] [[PubMed](#)]
62. Bradley, M.M.; Miccoli, L.; Escrig, M.A.; Lang, P.J. The pupil as a measure of emotional arousal and autonomic activation. *Psychophysiology* **2008**, *45*, 602–607. [[CrossRef](#)] [[PubMed](#)]
63. Rieger, G.; Savin-Williams, R.C. The Eyes Have It: Sex and Sexual Orientation Differences in Pupil Dilation Patterns. *PLoS ONE* **2012**, *7*, e40256. [[CrossRef](#)] [[PubMed](#)]
64. Tombs, S.; Silverman, I. Pupillometry: A sexual selection approach. *Evol. Hum. Behav.* **2004**, *25*, 221–228. [[CrossRef](#)]
65. Pronk, T.M.; Bogaers, R.I.; Verheijen, M.S.; Slegers, W.W. Pupil Size Predicts Partner Choices in Online Dating. *Soc. Cogn.* **2021**, *39*, 773–786. [[CrossRef](#)]
66. Anstis, S.M.; Mayhew, J.W.; Morley, T. The perception of where a face or television ‘portrait’ is looking. *Am. J. Psychol.* **1969**, *82*, 474–489. [[CrossRef](#)] [[PubMed](#)]
67. West, R.W. Perceived direction of gaze from eyes with dark vs. light irises. *Optomet. Vis. Sci.* **2011**, *88*, 303–311. [[CrossRef](#)]
68. Valtakari, N.V.; Hooge, I.T.C.; Viktorsson, C.; Nyström, P.; Falck-Ytter, T.; Hessels, R.S. Eye tracking in human interaction: Possibilities and limitations. *Behav. Res. Methods* **2021**, *53*, 1592–1608. [[CrossRef](#)] [[PubMed](#)]
69. Larsen, R.S.; Waters, J. Neuromodulatory Correlates of Pupil Dilation. *Front. Neural Circuits* **2018**, *12*, 21. [[CrossRef](#)] [[PubMed](#)]
70. Hömke, P.; Holler, J.; Levinson, S.C. Eye Blinking as Addressee Feedback in Face-To-Face Conversation. *Res. Lang. Soc. Interact.* **2017**, *50*, 54–70. [[CrossRef](#)]
71. van Gog, T.; Jarodzka, H. Eye Tracking as a Tool to Study and Enhance Cognitive and Metacognitive Processes in Computer-Based Learning Environments. In *International Handbook of Metacognition and Learning Technologies*; Azevedo, R., Aleven, V., Eds.; Springer: New York, NY, USA, 2013; pp. 143–156. [[CrossRef](#)]
72. Herwig, A.; Horstmann, G. Action–effect associations revealed by eye movements. *Psychonom. Bull. Rev.* **2011**, *18*, 531–537. [[CrossRef](#)]
73. Zeki, C.P. The importance of non-verbal communication in classroom management. *Proc. Soc. Behav. Sci.* **2009**, *1*, 1443–1449. [[CrossRef](#)]
74. Butt, M.N.; Sharif, M.M.; Muhammad, N.; Fanoos, A.; Ayesha, U. Eye contact as an efficient non-verbal teaching technique: A survey of teachers’ opinion. *Eur. J. Soc. Sci.* **2011**, *19*, 41–45.
75. Korthagen, F.A.J.; Attema-Noordewier, S.; Zwart, R.C. Teacher–student contact: Exploring a basic but complicated concept. *Teach. Teach. Educ.* **2014**, *40*, 22–32. [[CrossRef](#)]
76. Wubbels, T.; Brekelmans, M. Teacher–Students Relationships in the Classroom. In *Second International Handbook of Science Education*; Fraser, B.J., Tobin, K., McRobbie, C.J., Eds.; Springer: Dordrecht, The Netherlands, 2012; pp. 1241–1255. [[CrossRef](#)]
77. Sweeny, T.D.; Whitney, D. Perceiving Crowd Attention: Ensemble Perception of a Crowd’s Gaze. *Psychol. Sci.* **2014**, *25*, 1903–1913. [[CrossRef](#)]
78. Sung, G.; Feng, T.; Schneider, B. Learners learn more and instructors track better with real-time gaze sharing. *Proc. ACM Hum.-Comput. Interact.* **2021**, *5*, 1–23. [[CrossRef](#)]
79. Hirzle, T.; Sauter, M.; Wagner, T.; Rukzio, E.; Huckauf, A. Attention of Many Observers Visualized by Eye Movements. In *Workshop Proceedings of the 14th ACA Symposium on Eye Tracking Research Applications*; ACM: New York, NY, USA, 2022.
80. Dadds, M.R.; Perry, Y.; Hawes, D.J.; Merz, S.; Riddell, A.C.; Haines, D.J.; Solak, E.; Abeygunawardane, A.I. Attention to the eyes and fear-recognition deficits in child psychopathy. *Br. J. Psychiatr.* **2006**, *189*, 280–281. [[CrossRef](#)] [[PubMed](#)]
81. Chamberland, J.; Roy-Charland, A.; Perron, M.; Dickinson, J. Distinction between fear and surprise: An interpretation-independent test of the perceptual-attentional limitation hypothesis. *Soc. Neurosci.* **2017**, *12*, 751–768. [[CrossRef](#)]
82. Jack, R.E.; Garrod, O.G.B.; Schyns, P.G. Dynamic facial expressions of emotion transmit an evolving hierarchy of signals over time. *Curr. Biol.* **2014**, *24*, 187–192. [[CrossRef](#)] [[PubMed](#)]
83. West, R.W.; Van Veen, H.G. Gaze as Depicted in Vermeer’s Girl With a Pearl Earring. *J. Gen. Psychol.* **2007**, *134*, 313–328. [[CrossRef](#)] [[PubMed](#)]
84. Akechi, H.; Senju, A.; Uibo, H.; Kikuchi, Y.; Hasegawa, T.; Hietanen, J.K. Attention to eye contact in the West and East: Autonomic responses and evaluative ratings. *PLoS ONE* **2013**, *8*, e59312. [[CrossRef](#)] [[PubMed](#)]
85. Uono, S.; Hietanen, J.K. Eye contact perception in the west and east: A cross-cultural study. *PLoS ONE* **2015**, *10*, e0118094. [[CrossRef](#)] [[PubMed](#)]