I see you: insights into the neural and affective signatures of connectedness between parents and adolescents

Wever, M.C.M.

Chapter 6

Eyes on you: Ensuring empathic accuracy or signaling empathy?

Published as:
Wever, M.C.M., van Houtum, L.A.E.M., Janssen, L.H.C., Spruit, I.M., Tollenaar, M.S., Aan Het Rot, M., & Elzinga, B.M. (2022). Eyes on you: Ensuring empathic accuracy or signaling empathy? *International Journal of Psychology*. 57(6), 743-752. https://doi.org/10.1002/ijop.12862.

Link to preregistration prior to data analyses: https://osf.io/qxdv9/

ABSTRACT

The eye region is thought to play an important role in the ability to accurately infer others' feelings, or empathic accuracy (EA), which is an important skill for social interaction. However, most studies used static pictures, including only visual information, and knowledge about the contribution of the eye region to EA when visual information is presented together with verbal content is lacking. We therefore examined whether eye gazing contributes to EA during videos of emotional autobiographical stories including both visual and verbal content. One hundred seven perceivers watched videos of targets talking about positive and negative life events, and continuously rated the targets' feelings during the videos. Simultaneously, perceivers' eyes were tracked. After each video, perceivers reported on their feelings and the extent to which they empathized with and took the perspective of the targets. In contrast to studies using static pictures, we found that gazing to the eyes of targets during the videos did not significantly contribute to EA. At the same time, results on the association between the amount of gaze towards the eye region of targets and perceivers' state and trait empathy ratings suggest that eye gazing might signal empathy and social engagement to others.

Keywords: Empathic accuracy; Eye gaze; Empathic concern; Perspective taking; Social functioning.

INTRODUCTION

The ability to empathize with others is often considered a key ingredient for successful social interactions. However, the *accuracy* of inferring another's thoughts and feelings, also referred to as empathic accuracy (EA), is at least equally important (Ickes & Hodges, 2013; Zaki et al., 2008; Zaki et al., 2009). Several studies have emphasized the importance of the eye region for inferring the internal states of others (Baron-Cohen et al., 1997; Buchan et al., 2007; Eisenbarth & Alpers, 2011; Hall et al., 2010; Zaki et al., 2009). A task that emphasizes the importance of the eyes to infer the internal states of others is the reading the mind in the eyes task (Baron-Cohen et al., 1997). Numerous studies using this task have shown that a person's eye region contains sufficient information to identify complex mental states (Baron-Cohen et al., 1997). Moreover, the eye region automatically attracts and maintains attention, especially under emotional circumstances (Cowan et al., 2014) and people are generally found to gaze more towards the eyes of others compared to other facial features (e.g., mouth, nose or cheeks) (Buchan et al., 2007; Eisenbarth & Alpers, 2011; Hall et al., 2010).

Notably, however, EA in real life usually entails a mixture of visual and verbal information about the social situation involved. Nonetheless, most studies into the role of eye gaze in emotion recognition made use of static pictures including only visual information. This limits the generalizability to real life social interactions. Moreover, when the specific contribution of visual and verbal information of targets to the EA of perceivers is examined, verbal information has been found to contribute more to EA than visual information, but a combination of both produces the highest EA (Zaki et al., 2009). Furthermore, static pictures of clear facial emotional expressions do not accurately capture how our emotions are expressed in daily life, which can be much more subtle and ambiguous. So, while gazing to the eye region of others is beneficial under circumstances in which the informational source is limited to visual input, the added value of eye gazing when combined with verbal information is still unclear. So far, two studies have examined the association between eye gazing and trait empathy using a similar set of dynamic stimuli with both visual and verbal content (Cowan et al., 2014; Martínez-Velázquez et al., 2020). In these studies, people gazed more toward the eyes of others in emotional versus neutral video and the amount of eye gazing was positively associated with people's trait empathy levels. Building on these studies it is of interest to examine whether gazing toward the eyes of others also contributes to EA when visual information is presented together with verbal content.

Gazing to the eyes of others may be especially helpful in situations in which social cues are ambivalent. Happy faces are quickly recognized, and eye fixations are mostly directed to the mouth region, probably because a smile on the mouth is a clear and distinctive feature of happiness (Calvo et al., 2008; Eisenbarth & Alpers, 2011). Sad expressions, in contrast, generally include less distinctive facial features and people fixate more towards the eye region during these expressions, possibly to search for additional emotion cues (Bombari et al., 2013; Eisenbarth & Alpers, 2011). As such, looking at the eye region of others might contribute more strongly to EA during negative versus positive emotional situations. Besides, the eye region might be particularly informative when facial expressions are rather subtle or ambiguous (Baron Cohen et al., 1997; Vaidya et al., 2014). As facial expressions are generally less pronounced in less emotionally expressive persons, looking into the eye region of others might contribute more to EA when these "others" are less emotionally expressive.

In the present study, we aimed to determine whether gazing to the eyes of others contributes to EA when verbal content is present as well. Additionally, we examined whether this was dependent on the valence of the story content and targets' emotional expressivity. All measures and hypotheses of this study were preregistered at Open Science Framework prior to data analyses (https://osf.io/qxdv9/). We hypothesized that 1) perceivers who look more towards the eye region of targets show higher EA, 2) the amount of gaze towards the eyes of others is a stronger predictor of EA in negative versus positive videos, and 3) the amount of gaze towards the eyes of others is a stronger predictor of EA when targets are less emotionally expressive.

METHOD

Participants

Data were collected in the context of the RE-PAIR study: "Relations and Emotions in Parent-Adolescent Interaction Research". This study examines the relation between parent-adolescent interactions and adolescent mood. Families were eligible for inclusion if the adolescent and at least one of the parents were willing to participate and had a good command of the Dutch language. Further inclusion criteria were applied to the adolescents and can be found in Supplement S6.1. There were no additional in- or exclusion criteria for the parents.

The present study includes data of the parents of adolescents without psychopathology (n = 150); subsequently referred to as participants or as *perceivers*. Data of 43 participants were excluded: Complete task data of five participants was lost due to technical problems with the task and gaze data of 38 participants were missing due to unsuccessful calibration of the eye tracker. Reasons for the failure of calibration procedures were sight deficiencies, participants wearing glasses, or participants having light-colored eyes, which are all confirmed in prior studies to influence gaze data quality (Kammerer, 2009; Nyström et al., 2013). Of note, the EA task was embedded in a larger study for which the in- and exclusion criteria were not explicitly tailored to inclusion for eye tracking.

For 13 perceivers eye gaze data were missing on one or more videos (42 videos in total, range: 1-8) and 11 perceivers had <70% valid gaze data on one or more videos (37 videos in total, range: 1-8). In addition, 17 perceivers of the final sample were missing continuous EA ratings on one or more videos (37 videos in total, range: 1-5) due to technical problems during the task or inadequate use of the dial. This resulted in a final sample of 107 perceivers with 981 videos in total for the analyses (out of 1070, 8.3% missing data), including 48 males (45%, M_{age} = 50 years, *SD* = 5.97) and 59 females (55%, M_{age} = 47 years, *SD* = 4.75). The final sample (*n* = 107) was representative for the total number of participants that performed the EA task (*n* = 150) as they did not significantly differ on age, gender, trait empathy, autism spectrum traits, and intellectual functioning.

The study was approved by the medical ethical committee of the Leiden University Medical Centre (LUMC) (P17.241) and was performed in accordance with the declaration of Helsinki and the Dutch Medical Research Involving Human Subjects Act (WMO). All participants provided written informed consent at the start of all study visits and were blind to the hypotheses of the present study.

Procedure

Families were recruited via public places and social media. Parents and adolescents were briefed about the study and underwent a comprehensive telephone screening during which family circumstances and verbal informed consent were discussed. When found eligible for participation, families were invited for a lab visit to the Leiden University Treatment and Expertise Centre (LUBEC) in Leiden. Two weeks prior to the appointment participants were asked to fill out an online questionnaire battery that included questions about demographics and clinical and cognitive constructs (see Measures and materials). During the lab visit, families performed parent-adolescent interaction tasks and filled out additional questionnaires, parents were screened for psychopathology, and intellectual abilities were assessed. Furthermore, parents performed the EA task while eye tracking measures were taken, which is the focus of the present study.[1]

Measures and materials

Empathic accuracy task

Similar to the English task (Zaki et al., 2008), the Dutch version of the EA task developed by aan het Rot and Hogenelst (2014) includes dynamic stimuli of various target people who are narrating both positive/happy (e.g., celebrating a birthday with friends) and negative/sad (e.g.,

[1] After the lab visit, families were (a) asked to fill out ecological momentary assessments for 14 consecutive days on their smartphones, using a mobile app to assess affect and parent-adolescent interactions in daily life, and (b) invited for an MRI session on a separate day. Data derived from these parts of the RE-PAIR study were reported elsewhere.

a friend died of a brain hemorrhage) emotional autobiographical stories. Within 30 minutes after the stories of the targets were videotaped, the targets watched their personal recordings and continuously rated how they felt in their videos by using a dial. The dial included a Likert scale ranging from 1 (*extremely negative*) to 9 (*extremely positive*). Additional information about task development can be found in aan het Rot and Hogenelst (2014).

The targets varied in their self-reported emotional expressivity as assessed with the Berkeley Expressivity Questionnaire (BEQ; Gross and John (1997)). The BEQ consists of 16 items that are answered on a Likert scale from 1 (*strongly disagree*) to 7 (*strongly agree*). Mean BEQ scores were calculated by averaging all items and ranged in the present target sample from 3.50 – 5.97. Higher scores represent higher emotional expressivity. Prior studies reported a significant positive association between BEQ scores and EA (aan het Rot & Hogenelst, 2014; Zaki et al., 2008). Since the original item of targets on the BEQ could not be retrieved the reliability of the scale in the current sample could not be computed. Yet, previous studies found good validity and reliability for the instrument (α = .86) (Kupper et al., 2020), and there is no reason to expect any differences between these studies and the use of the BEQ in the present sample.

The present study includes a shortened version of the EA task with a duration of ±25 minutes instead of the original 50 minutes. This was done due to time constraints as the task was part of a larger study protocol. The task included a subset of 10 videos, consisting of five positive and five negative autobiographical stories from six distinct targets (three males, three females) that derived from a pilot study in which we tested the feasibility of the EA task in combination with eye tracking (Supplement S6.2).

The participants in the present study (perceivers) were asked to watch all 10 videos and were instructed to imagine that they were sitting on the other side of the table of the targets while listening to their autobiographical stories. At the same time, they continuously rated how they thought the target was feeling while narrating, using the same dial as the targets used to rate their feelings. Videos were semi-randomly presented, with no more than two positive or negative videos and no more than two videos with a target of the same gender in a row. Prior to the start of the task perceivers were presented with a test trial in which the test leader checked correct use of the dial. Prior to the start of each video perceivers were asked to put the dial back to “neutral” to start each trial from the same position. All procedures are in line with previous studies using this task (aan het Rot & Hogenelst, 2014).

A new addition was that after each video the perceivers were asked to report on how well they were able to empathize with (i.e., *state* empathic concern) and put themselves in the shoes (i.e., *state* perspective taking) of the target. Also, perceivers rated how happy, sad, relaxed, and irritated they felt after each video. All questions were rated on a Likert scale ranging from 1 (*not at all)* to 7 (*very much*).

Stimulus presentation and simultaneous eye movement recordings were conducted using E-Prime 2.0 software with the E-Prime Extension for Tobii package (Psychology Software Tools, Pittsburgh, PA, United States). The screen resolution was 1920x1080 pixels and videos were presented on the screen in 960x540 pixels.

Eye tracking

Eye movements were recorded with a portable Tobii Pro X3-120 eye tracker sampling at 120 Hz. Prior to the start of the task perceivers were asked to place their head in a chin rest to prevent head movement during the recording and the distance to the screen was set at 60 cm. Perceivers' eyes were calibrated using a 9-point calibration grid and calibration results were visually inspected and accepted if quality was approved. In case of missing calibration points or poor calibration quality, the procedure was repeated for a maximum of three attempts after which the quality was unlikely to further improve. The EA task started directly after the calibration procedure and gaze data was recorded until the task was finished.

Trait empathy

To assess trait empathy perceivers filled out the empathic concern (EC) and perspective taking (PT) subscales of the interpersonal reactivity index (IRI) prior to the start of the lab visit (Davis, 1980; De Corte et al., 2007). EC includes the reported tendency to experience feelings of sympathy and compassion for unfortunate others and PT includes the reported tendency to spontaneously adopt the psychological point of view of others in everyday life. Both subscales include 7 items and are answered on a Likert scale ranging from 0 (*does not describe me well*) to 4 (*describes me very well*). Sum scores of each subscale were calculated by adding up the items (range in the present sample was 7–28 for EC and 6–27 for PT). Higher scores represent higher trait empathy levels. The validity and reliability of the Dutch IRI have been established (De Corte et al., 2007) and the internal consistencies of the subscales in the present sample were acceptable ($\alpha = 0.75$ for both).

Intellectual functioning

Intellectual functioning was assessed with two subtests of the Dutch Wechsler Adult Intelligence Scale IV (WAIS-IV-NL; Wechsler (2012)): Block design (perceptual organization skills) and vocabulary (verbal skills). Individual raw scores were translated into norm scores based on age and were averaged to calculate the estimated intellectual functioning measure per individual. This measure was included as covariate in the analyses to control for individual differences in intellectual functioning. Validity of this subtest dyad with the original full scale IQ has been established (Girard et al., 2015).

Data analyses

Preprocessing

Preprocessing of the raw data from the EA task was similar to aan het Rot and Hogenelst (2014), with raw continuous ratings from perceivers and targets being preprocessed into an EA measure per video in SAS 9.3 for Windows (SAS, Cary, NC). For data reduction purposes, ratings from perceivers and targets were averaged across five-second periods. The last five seconds of all ratings were discarded, because it included the return of the dial to the "neutral" position before the end of each video. Subsequently, first-order autocorrelations were removed from the continuous ratings using the Yule-Walker method. For each video we correlated perceiver ratings of the target's feelings and target ratings of their own feelings, resulting in a correlation coefficient r that defined the perceiver's raw EA score per video. Raw EA scores underwent a Fisher z transformation prior to further analyses.

See Supplement S6.3 for more details on the preprocessing of raw eye tracking data into measures of eye gaze per perceiver per video. The primary eye gaze measure is the percentage of dwell time within the defined areas of interest (AOIs; i.e., eyes, mouth, and face as a whole) per video, as part of the total video duration, in which dwell time is defined as the total amount of time spent looking within an AOI and includes all types of eye movements. The percentage of dwell time within the face and mouth AOI were described to identify to what extent perceivers gazed towards the face and mouth of the targets in addition to their eye region. Dwell time is interpreted as the level of interest in an AOI, with greater dwell times indicating greater levels of interest.

Statistical Analyses

Means and standard deviations of the EA task and the self-report ratings per video and valence category (i.e., positive or negative) were calculated. In addition, the average percentage of dwell time for each AOI (i.e., eyes, mouth, and face) per valence category and video were assessed.

The effects of our hypothesized predictors on EA were tested in R-3.6.1 (R Core Team, 2013), using generalized linear mixed regression models with a multi-level, within-subject design. We used lme4 for multilevel analyses with maximum likelihood (Bates et al., 2012) and ggplot2 for figures (Wickham et al., 2016). The dependent variable EA has been repeatedly measured and EA observations per video (level 1) were nested within perceivers (level 2). Predictor variables that act upon the perceiver-level were the percentage of dwell time within the eye region and trait EC and PT. Predictor variables that act upon the target-level (level 1) were target expressivity and valence of the videos.

First, we ran correlations between all predictor variables (i.e., percentage of dwell time within the eye region, target expressivity, valence, and trait EC and PT) and the outcome variable EA. Thereafter, we tested the validity of the task by assessing the influence of target expressivity, valence, and trait EC and PT on EA using generalized linear mixed regression models.

Subsequently, we tested our main hypotheses about the influence of the percentage of dwell time within the eye region of targets on EA (hypothesis 1) and whether this interacts with the valence of the videos (hypothesis 2) or target expressivity (hypothesis 3) in two separate interactions. Exploratively, we also tested a three-way interaction between the percentage of dwell time within the eye region of targets, target expressivity, and valence on EA. In case of significant interactions, we broke down the interaction into simple contrasts using Bonferroni-corrected *post hoc* tests.

To check whether results were not driven by differences in age, gender, and intellectual functioning of perceivers, we performed additional analyses in which we statistically controlled for these variables. Significance was set at $p < .05$ (two-tailed) and Cohen's d effect sizes were calculated for significant effects.

RESULTS

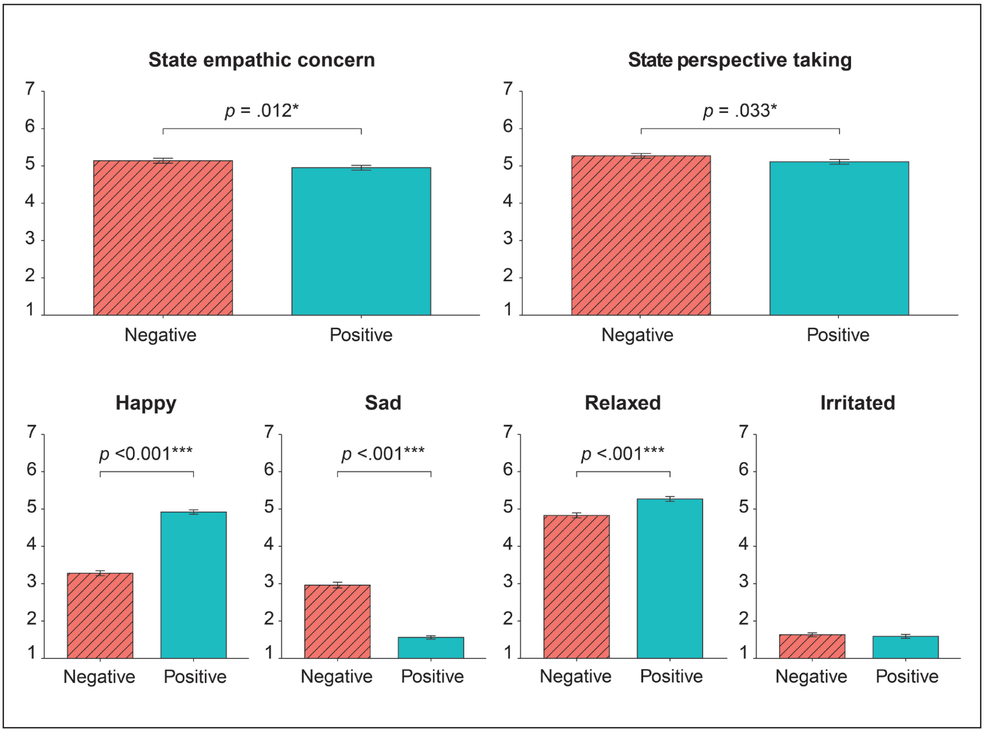
Task descriptives

See Table 6.1 for demographic and clinical characteristics. Data derived from individual state empathy ratings after each video revealed that perceivers empathize more with and were better able to take the perspective of the targets during negative versus positive videos ($B = 0.19$, SE $= 0.07$, $t(872.39) = 2.52$, $p = .012$, $d = 0.17$ and $B = 0.16$, SE $= 0.07$, $t(871.99) = 2.13$, $p = .033$, $d = 0.14$, respectively). In addition, they felt significantly more sad after negative videos ($B = 1.40$, SE $= 0.07$, $t(874.21) = 18.89$, $p < .001$, $d = 1.28$) and happier ($B = 1.65$, SE $= 0.07$, $t(870.35) = 22.13$, $p < .001$, $d = 1.50$) and relaxed ($B = 0.45$, SE $= 0.06$, $t(870.75) = 7.17$, $p < .001$, $d = 0.49$) after positive videos. There was no significant difference in irritability between positive and negative videos ($p = .366$; Figure 6.1). In addition, perceivers who reported higher trait EC and PT were also better able to empathize with ($B = 0.05$, SE $= 0.02$, $t(101.98) = 2.86$, $p = .005$, $d = 0.57$) and take the perspective of targets ($B = 0.06$, SE $= 0.02$, $t(105.03) = 3.05$, $p = .003$, $d = 0.60$) based on the state empathy ratings per perceiver after each video (see Supplement S6.4 and S6.5).

Table 6.1 Demographics and clinical data

Mean (SD)	Total *n* = 107	Males *n* = 48	Females *n* = 59	Gender differences [1]
Age	48 (5.50)	50 (5.97)	47 (4.75)	.005
Autism spectrum traits [2]	54.25 (10.48)	57.92 (11.01)	51.27 (9.08)	<.001
Trait empathy [3]				
Empathic concern	18.08 (4.92)	15.79 (4.29)	19.95 (4.63)	<.001
Perspective taking	17.27 (4.54)	15.94 (4.67)	18.36 (4.15)	.006
WAIS-IV [4]				
Block design	10.92 (3.16)	11.73 (3.09)	10.23 (3.09)	.020
Vocabulary	11.68 (2.49)	11.73 (2.60)	11.64 (2.60)	.857
Average score	11.30 (2.15)	11.73 (2.28)	10.93 (1.99)	.071

Note. WAIS-IV, Wechsler Adult Intelligence Scale IV (Wechsler, 2012). [1] *p*-values were obtained using independent samples *t*-test comparisons between males and females. [2] Autism spectrum traits were assessed with the 28-item version of the Autism-spectrum Quotient (AQ-short; Hoekstra et al. (2011)). [3] Trait empathy was assessed with the Interpersonal Reactivity Index (IRI; De Corte et al. (2007)). [4] Data on intellectual functioning (WAIS-IV) were missing or unreliable for 11 participants resulting in $n = 96$ for this measure.



6

Figure 6.1 *Mean individual ratings of perceivers, plotted for negative and positive videos, rated on a Likert scale from 1 (not at all) to 7 (very much) after each video. Significance was tested with generalized linear mixed regression model analyses. Error bars represent standard error of the mean. Significant p-values <.05 were indicated by *, p <.01 by **, and p <.001 by ***.*

The mean raw *r* between perceivers' and targets' EA scores was 0.53 and did not differ between males and females. We ran generalized linear mixed regression model analyses in which we assessed the influence of valence, target expressivity, and trait EC and PT on EA. As expected, but in contrast to the impact of valence on perceivers' state empathy levels, perceivers were less empathically accurate during negative versus positive videos (B = -0.46, SE = 0.06, $t(881.9) = -8.25$, $p < .001$, $d = 0.56$). Target expressivity and trait EC and PT of perceivers were not significantly associated with perceivers' EA (all $p \geq .796$). All outcomes remained significant after controlling for age, gender, or intellectual abilities of perceivers in separate analyses.

On average, perceivers gazed for 85.7% (SD = 8.65%) of the total duration of the videos towards the faces of targets, indicating that the targets' faces substantially attracted and maintained perceivers' attention. In addition, perceivers gazed on average for 33.38% (SD = 18.49%) of the total duration of the videos to the eye region of the targets. There was no significant difference between males and females in the percentage of dwell time within the eye region of the targets. Perceivers gazed more into the eyes of others during negative versus positive videos (B = 3.77, SE = 0.61, *t*(873.75) = -6.15, $p < .001$, *d* = 0.42). In addition, perceivers with higher trait EC and PT gazed significantly more into the eyes of others (trait EC: *B* = 0.73, SE = 0.36, *t*(104.96) = 2.02, *p* = .046, *d* = 0.40; trait PT: *B* = 0.79, SE = 0.39, *t* (105.65) = 2.03, *p* = .045, *d* = 0.40), independent of the emotional valence of the videos. In line with this, perceivers who gazed more within the eye region of targets during a video reported to empathize more with and were better able to take the perspective of the targets narrating the autobiographical story on a state level (*B* = 0.70, SE = 0.28, *t*(906.47) = 2.47, *p* = .014, *d* = 0.16 and *B* = 0.67, SE = 0.29, *t*(905) = 2.33, *p* = .020, *d* = 0.16, respectively). Target expressivity was not significantly associated with perceivers' dwell time within the eye region of the targets (*p* = .571). On average, there was 9.46% (SD = 6.53%) missing gaze data during which participants gazed outside of the computer screen. The amount of missing gaze data was not dependent on the presentation order of the videos in the task. For more details on missing gaze data over the course of the task, see Supplement S6.9.

In addition to the eye region, perceivers gazed on average for 15.79% (*SD* = 15.14%) of the total duration of the videos to the mouth of the targets. Male and female perceivers did not differ significantly in the percentage of dwell time to the mouth of the targets. In addition, valence and trait EC and PT were not significantly associated with the percentage of dwell time to the mouth of the targets. However, we found a positive association between emotional expressivity of targets and the percentage of dwell time of perceivers to the mouth of targets in the videos, with perceivers gazing more to the mouth region of more (compared to less) expressive targets (*B* = 0.89, SE = 0.21, *t*(873.56) = 4.25, $p < .001$, *d* = 0.04, Supplement S6.4-S6.8.

Effects of gazing to the eyes on empathic accuracy

With regard to the main focus of our study (hypothesis 1), the percentage of dwell time within the eye region of targets was not significantly related to perceivers' EA (*p* = .146). In addition, there was no significant interaction between the percentage of dwell time within the eye region of targets and targets' emotional expressivity on perceivers' EA (hypothesis 3) (*p* = .416). We did find a significant interaction between the percentage of dwell time within the eye region of targets and the emotional valence of the videos on perceivers' EA (hypothesis 2) (*B* = -0.01, SE = 0.002, *t*(892.78) = -3.33, $p < .001$, *d* = 0.22), although in opposite direction. In contrast with our expectations, there was no significant association between gazing to the eye region of targets and EA in negative videos, however, perceivers that gazed *more* into the eye region of the targets during positive videos were somewhat *less* empathically accurate (Figure 6.2, Supplement S6.10).

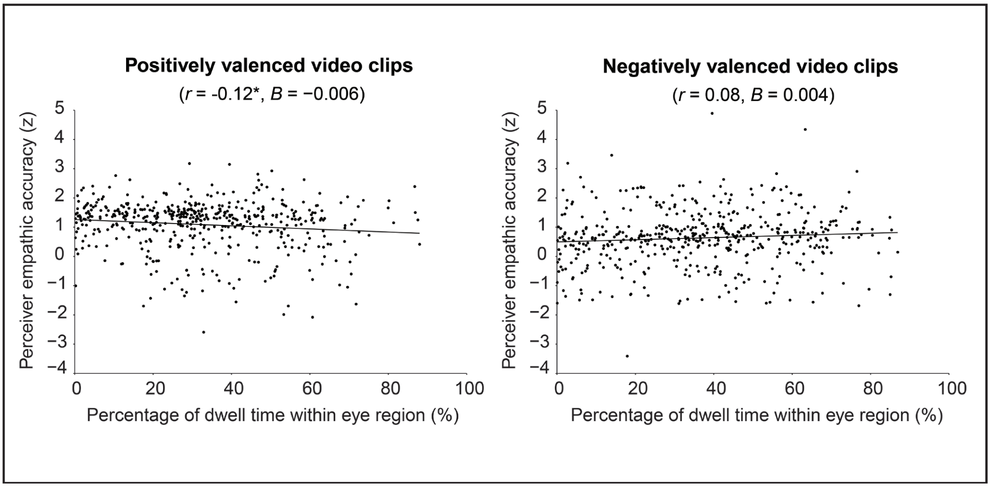


Figure 6.2 *Associations between the percentage of dwell time within the eye region of targets and perceivers' EA in positive and negative videos. Significant p-values <.05 were indicated by *, p <.01 by **, and p <.001 by ***.*

Explorative analyses

The finding that perceivers who gazed more into the eye region of targets during positive videos were less empathically accurate raised the question whether perceivers instead gazed more to the mouth during these videos. Therefore, we additionally explored whether valence also interacted with the percentage of dwell time to the mouth of the targets. We examined the interaction between the percentage of dwell time to the mouth and emotional valence of the videos on perceivers' EA, but this interaction was non-significant (p = .063). Also, there was no significant interaction between the percentage of dwell time to the mouth of targets and targets' emotional expressivity on perceivers' EA (p = .752) or between the percentage of dwell time to the mouth of targets and perceivers' EA in general (p = .860).

6

DISCUSSION

The present study used a paradigm with high ecological validity to examine whether gazing to the eyes of others contributes to EA during videos of emotionally valenced target stories in which verbal information was also available. First, gazing to the eyes of others did not significantly contribute to EA. Second, however, the emotional valence of the stories did moderate the relation between gazing to the eyes of others and EA. Perceivers who gazed more to the eye region of others during positive target stories were less empathically accurate, whereas this was not found during negative target stories. Third, targets' emotional expressivity was not significantly related to perceivers' EA, nor did it moderate the relation between gazing to the eyes of others and perceivers' EA.

In contrast to our hypotheses we found that perceivers who gazed more towards the eye regions of targets were not more empathically accurate. Moreover, perceivers who gazed less towards the eyes of others during positive videos were even more empathically accurate. Although the importance of the eye region has consistently been demonstrated in studies that solely convey visual input, our findings indicate that the eye region seems to be less informative when visual input is presented in co-occurrence with verbal information. While prior studies have greatly contributed to our basic understanding of the role of the eye region in social interactions, the current results emphasize the importance of also studying such processes in more ecologically valid settings since conclusions can deviate in important ways.

Our results did not show an effect of target expressivity on perceivers' EA scores, nor did it moderate the relation between gazing to the eyes of targets and perceivers' EA. This was not in line with our hypotheses and prior studies using the Dutch EA task, but this might be due to differences in methodology. In the current, shortened, version of the EA task we included six out of 11 targets, which considerably decreased the diversity of target expressivity in the present study. In addition, the BEQ mainly focuses on emotional expressivity in the face of targets, while perceivers also receive verbally expressive informational cues of the targets to base their EA on. It is possible that targets who report to have less expressive faces could still have an expressive tone of voice, which might have revealed information about their internal state.

An unexpected finding was that perceivers who considered themselves more empathic, both at trait (EC and PT scales of the IRI) and state level (individual ratings after each video), gazed more towards the eye region of others. In this light, gazing to others' eyes might be a way to express empathy to others, rather than (only) collect (additional) socio-emotional information about others' internal state. This dovetails with the results of Cowan et al. (2014) and Martínez-Velázquez et al. (2020), who interpreted the increased gazing towards the eye region of others as enhanced social engagement. Moreover, looking at the eyes of a conversational partner while listening was found to signal interest and affiliation (Breil & Böckler, 2021).

We found that perceivers gazed more towards the eye region of others during negative versus positive videos. In addition, individual ratings of perceivers' state empathy showed that they were better able to empathize with and take the perspective of targets in negative versus positive videos. This is in accordance with the possible signaling function of eye gaze, suggesting that people might have a natural tendency to empathize with and gaze more to the eyes of others during negative versus positive emotional situations. This effect may have been emphasized by the stimuli duration (±1 – 2 minutes), as empathic feelings are particularly induced after prolonged presentation (Regenbogen et al., 2012). It is of note, however, that participants were instead less empathically accurate during negative versus positive videos, pointing to the distinct impact of the valence of the videos on participants' feelings of empathy versus their levels of empathic accuracy. Hence, feeling empathy and being empathically accurate in inferencing what others might feel is not the same.

A signaling function of eye gaze has been previously mentioned in the literature (Cowan et al., 2014; Kobayashi & Hashiya, 2011; Mason et al., 2005), although empirical evidence was lacking. Kobayashi and Hashiya (2011), for example, introduced the "gaze-grooming" hypothesis, stating that gaze has evolved into a contact-free, social grooming function in humans to form and maintain social bonds. Our results are in line with this "gaze-grooming" hypothesis and the various target stories deriving from distinct targets show empirical evidence for the generalizability of this signaling function of gaze to a variety of social situations.

Strengths and limitations

This study uniquely examined to what extent gazing at the eye region of others contributes to participants' EA under ecologically valid circumstances. The methodological design of the EA task not only allows for a corresponding assessment of the feelings of both perceiver and target in positive and negative situations, but also incorporates the assessment of fluctuations in their affect over time. Furthermore, the novel addition of individual ratings about perceivers' affect and state empathy after each video informed us on how participants subjectively experienced the emotional target stories and gives additional insight in the validity of the task.

While the richness of the dynamic stimuli, including both verbal and non-verbal information, are a major advantage of the present study, future studies could focus on the individual contribution of the verbal and visual content to EA (Zaki et al., 2009). As perceivers were presented with videos of unknown targets, they were well aware that they were not involved in an actual bidirectional conversation. This may have lowered their motivation to be empathically accurate and may have affected our findings. Related to this, the videos do not mimic bidirectional interactions, but rather mimic listening to a monologue. It is important to mention that these are two different types of interactions that occur under different circumstances. As the EA task more closely mimics the latter, our findings are probably most generalizable to closely resembling situations in real life, such as (mental) health settings in which practitioners are listening to personal stories of their clients. Furthermore, participants were placed in a chin rest while watching the videos to limit head motion. Although they reported low irritability during the task and EA levels were comparable to prior studies using the EA task, it is possible that they experienced the chin rest as unpleasant, which might have affected their performances. Lastly, it is of note that the participants in this study are adults aged between 35 and 64 years (M_{age} = 48; SD = 5.50) and the results of the present study need to be interpreted in the context of this age group.

Conclusion

While prior studies have shown the importance of the eye region for inferring others' feelings when only visual information is available, our results show that gazing to the eyes of others may not contribute to EA when both visual and verbal information is available. In addition, gazing to the eyes of others seem to be a way to express empathy and social engagement to others. In other words, our results inform us on the role of eye gazing during social interactions and shed light on a possible signaling function of eye gazing to sympathize or empathize with our conversational partners.

This outcome, compared to that obtained using less ecologically valid paradigms, emphasizes the importance of studying how individuals perceive others in social settings that closely mimic real life. Our findings enrich the field of social sciences in several ways and implicate that we need to be very careful in translating findings from basic science to the complex realm of daily life. On a theoretical level, there is a clear need to better understand the factors that contribute to EA in daily life, as our data seem to suggest that gazing to the eyes is not a substantial source of information in our daily conversations. At the methodological level, these results make us aware of the way methodological differences between studies give rise to diverging outcomes and that a combination of both basic experiments and designs including more ecologically valid measures is needed to better understand social interactions. Lastly, our findings may have implications at the practical level for communication between people in general, and might be of particular relevance for health care practitioners in medical or therapeutic settings. Signaling their empathy and emotional engagement by gazing into the eyes of their clients, especially when listening to their (emotionally valenced) personal stories, might be particularly helpful in favoring the quality of the therapeutic relationship.

SUPPLEMENTARY MATERIALS

SUPPLEMENT S6.1

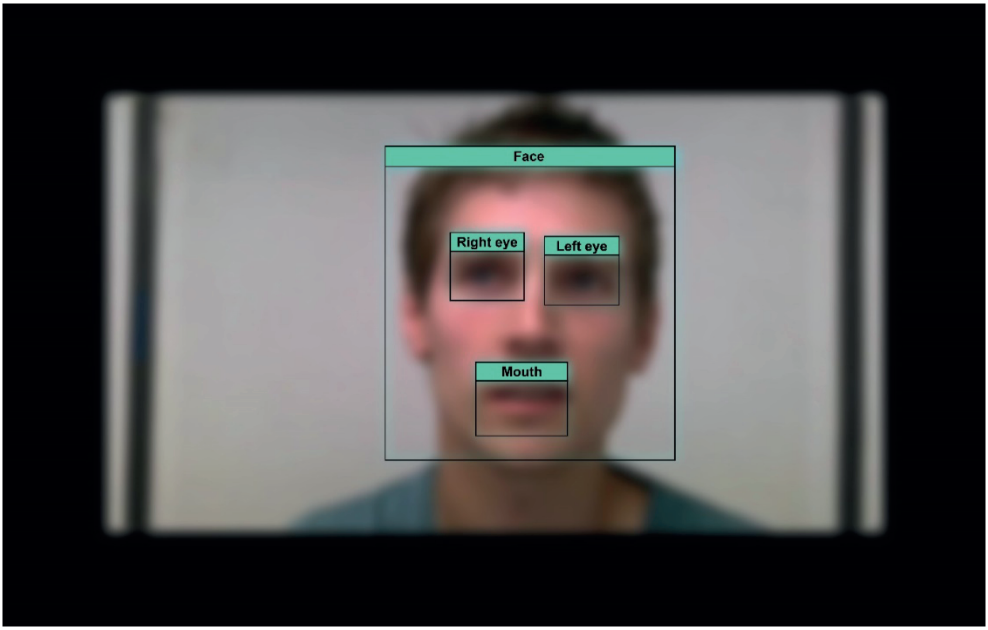
Inclusion criteria were applied to the adolescent and included being aged between 11 and 17 years, and living with at least one of their parents. All adolescents were screened on current and lifetime psychopathology with the Kiddie-Schedule for Affective Disorders and Schizophrenia-Present and Lifetime Version (K-SADS PL; Kaufman et al. (1996)). Adolescents without psychopathology and their parents were included in the study if the adolescents did not meet criteria for any (neuro)psychiatric disorder in the two years leading up to the study, and had no lifetime diagnoses of MDD/dysthymia. Adolescents with psychopathology and their parents were included if the adolescent met criteria for a current, primary diagnosis of MDD/dysthymia according to the K-SADS.

SUPPLEMENT S6.2

The present study includes a shortened version of the EA task with an approximate duration of 25 minutes instead of the 50 minutes of the original version of the task. The subset of videos was derived from a pilot study in a student sample ($n = 20$) in which we tested the feasibility of the already existing EA task in combination with an eye tracking set-up. Videos were selected based on their feasibility for eye tracking purposes (i.e., videos in which targets were excessively moving were excluded) and sufficient variety in EA ratings per video in the pilot study (range mean EA between 0.35-0.85; $SD > 0.30$). Also, we tested two versions of the task in which one version contained the presentation of the 9-point Likert scale of the dial below the video on the screen (similar to the original task), while the other version did not include the presentation of this scale on the screen. The data of the pilot indicated that participants often looked at the scale on the screen instead of the video, which was not desirable for the eye tracking measures. We therefore decided to use the version without the scale on the screen for the present study.

SUPPLEMENT S6.3

A customized MATLAB script (MathWorks, Inc., Natick, MA, version 9.5) was used to pre-process raw eye tracking data into measures of eye gaze per perceiver per video. Raw gaze data of at least one eye was used to calculate information of gaze position and duration. Furthermore, validity of gaze data was calculated as the percentage successfully recorded eye tracking data per video as an estimate of data quality. Individual videos of which the validity was below 70% were excluded from analyses. In order to follow the natural movement of the targets in the videos, dynamically moving areas of interest (AOI) were created around the left eye, right eye, mouth, and face as a whole of all individual targets using MATLABs cascade object detector, which uses the algorithm of Viola and Jones (2001) for face and facial feature detection. More specifically, for each frame of each video, this algorithm outputted rectangular AOIs encompassing the left eye, right eye, mouth, and face (see below). Outlier removal, smoothing, and interpolation was performed on the AOIs thereafter, to correct any incorrectly identified AOIs due to movement or blinking of the target in the video. The gaze data within the right and left eye AOIs were corrected for overlap and combined into a single AOI for the eye region. The screenshot of the target person presented below is blurred due to privacy reasons.



SUPPLEMENT S6.4

Empathic accuracy and gaze data for positive and negative videos

Mean (SD)	All videos *n* = 981	Positive videos *n* = 486	Negative videos *n* = 495	Differences in valence [1] *t*	*p*
EA (Fisher z transformed)					
All	0.84 (0.40)	1.09 (0.45)	0.63 (0.51)	8.25	<.001
Males	0.79 (0.44)	1.03 (0.51)	0.53 (0.53)	5.52	<.001
Females	0.93 (0.35)	1.15 (0.40)	0.70 (0.49)	6.13	<.001
Dwell time face AOI, %					
All	85.70 (8.65)	85.34 (8.79)	86.05 (8.74)	-2.43	.015
Males	85.56 (9.04)	84.97 (9.29)	86.15 (9.02)	-2.67	.007
Females	85.81 (8.40)	85.64 (8.42)	85.96 (8.57)	-0.91	.363
Dwell time eyes AOI, %					
All	33.38 (18.49)	31.39 (17.76)	35.39 (19.67)	-6.15	<.001
Males	32.60 (20.69)	30.60 (20.03)	34.59 (21.73)	-4.49	<.001
Females	34.02 (16.65)	32.02 (15.82)	36.04 (17.99)	-4.22	<.001
Dwell time mouth AOI, %					
All	15.79 (15.14)	15.58 (14.86)	16.02 (15.56)	-1.18	.240
Males	17.40 (17.00)	17.08 (16.74)	17.74 (17.35)	-0.57	.569
Females	14.48 (13.45)	14.36 (13.15)	14.62 (13.93)	-1.07	.287
Individual state empathy and affect ratings after each video					
Empathic concern	5.05 (1.44)	4.95 (1.42)	5.14 (1.47)	-2.52	.012
Perspective taking	5.19 (1.41)	5.11 (1.42)	5.27 (1.39)	-2.13	.033
Happy	4.11 (1.60)	4.92 (1.30)	3.28 (1.45)	22.13	<.001
Sad	2.25 (1.56)	1.56 (1.00)	2.96 (1.71)	-18.89	<.001
Irritated	1.61 (1.13)	1.59 (1.14)	1.63 (1.13)	-0.91	.366
Relaxed	5.06 (1.45)	5.27 (1.39)	4.83 (1.47)	7.17	<.001

Note. AOI, area of interest; EA, empathic accuracy; SD, standard deviation. [1] Differences in valence were calculated by generalized linear mixed regression models and were tested without covariates in the model.

SUPPLEMENT S6.5

EA task data perceivers per video

Videos	Duration (s)	Valence	*n*	Perceiver ratings per video, mean (SD)						
				EA Fisher *z* transformed	Empathic concern	Perspective taking	Happiness	Sadness	Relaxed	Irritated
1	116	N	98	1.15 (1.33)	5.59 (1.38)	5.72 (1.31)	2.60 (1.33)	3.61 (1.78)	4.57 (1.60)	1.46 (0.98)
2	83	P	97	1.17 (0.59)	4.59 (1.43)	4.75 (1.51)	4.60 (1.27)	1.59 (1.06)	5.32 (1.26)	1.58 (1.13)
3	107	N	95	0.56 (1.34)	4.66 (1.56)	4.91 (1.50)	3.48 (1.30)	2.72 (1.48)	4.97 (1.35)	1.80 (1.37)
4	104	N	96	0.30 (0.62)	4.74 (1.39)	4.84 (1.33)	3.93 (1.34)	2.36 (1.40)	4.94 (1.47)	1.61 (0.96)
5	155	N	101	0.66 (0.46)	5.68 (1.27)	5.69 (1.22)	2.55 (1.35)	3.71 (1.82)	4.61 (1.48)	1.74 (1.35)
6	103	P	99	0.55 (0.90)	4.87 (1.46)	5.07 (1.47)	4.78 (1.37)	1.78 (1.15)	5.29 (1.28)	1.86 (1.30)
7	102	P	99	1.41 (1.09)	4.62 (1.49)	4.79 (1.43)	4.55 (1.43)	1.70 (1.08)	5.19 (1.46)	1.81 (1.33)
8	105	P	100	1.11 (0.80)	4.95 (1.31)	5.14 (1.33)	5.06 (1.13)	1.49 (0.89)	5.20 (1.44)	1.45 (1.06)
9	121	N	96	0.48 (0.41)	4.99 (1.44)	5.16 (1.34)	3.88 (1.28)	2.32 (1.46)	5.09 (1.38)	1.54 (0.92)
10	109	P	100	1.20 (0.49)	5.72 (1.08)	5.80 (1.11)	5.62 (0.97)	1.26 (0.68)	5.37 (1.52)	1.24 (0.62)

Note. EA = Empathic accuracy; N = negative video; P = positive video; S = Seconds; SD = Standard deviation.

6

SUPPLEMENT S6.6

Generalized linear mixed regression model to assess the influence of valence, target expressivity, and empathic concern and perspective taking of perceivers on the level of EA (Fisher z transformed)

	B	**SE**	**df**	*t*	***p***
Model 1					
Intercept[1]	0.63	0.05	235.6	13.22	<.001
$Valence_{positive}$	0.46	0.06	881.9	8.25	<.001
Model 2					
Intercept[1]	-0.92	0.16	959.7	5.75	<.001
BEQ	-0.01	0.04	883.7	-0.35	.730
Model 3					
Intercept[1]	0.81	0.15	104.4	5.48	<.001
Empathic concern	0.01	0.01	103.5	0.34	.733
Model 4					
Intercept[1]	0.82	0.16	106.8	5.30	<.001
Perspective taking	0.01	0.01	106.7	0.26	.796

[1] The intercept includes perceivers' level of empathic accuracy (Fisher *z* transformed) during negative autobiographical stories.

SUPPLEMENT S6.7

Perceivers' average percentage of dwell time per AOI per video

				% Dwell time per AOI, mean (SD)		
Videos	**Duration (s)**	**Valence**	***n***	**Eyes**	**Mouth**	**Face**
1	116	N	98	33.01 (21.96)	15.64 (16.33)	86.09 (9.46)
2	83	P	97	30.90 (19.96)	15.09 (15.53)	85.40 (10.02)
3	107	N	95	34.97 (21.72)	16.90 (16.17)	87.29 (8.78)
4	104	N	96	35.51 (20.67)	17.42 (16.68)	87.38 (8.41)
5	155	N	101	43.96 (22.30)	15.54 (14.99)	85.89 (8.95)
6	103	P	99	42.71 (22.23)	17.16 (17.26)	85.25 (8.50)
7	102	P	99	31.49 (17.74)	17.50 (14.84)	86.90 (8.43)
8	105	P	100	28.49 (16.50)	12.13 (14.00)	86.52 (8.74)
9	121	N	96	33.89 (19.76)	15.88 (16.91)	87.77 (8.35)
10	109	P	100	29.12 (17.96)	16.76 (16.51)	86.89 (8.01)

Note. AOI = area of interest; N = negative video; P = positive video; S = seconds; SD = Standard deviation.

6

SUPPLEMENT S6.8

EA task data targets per video

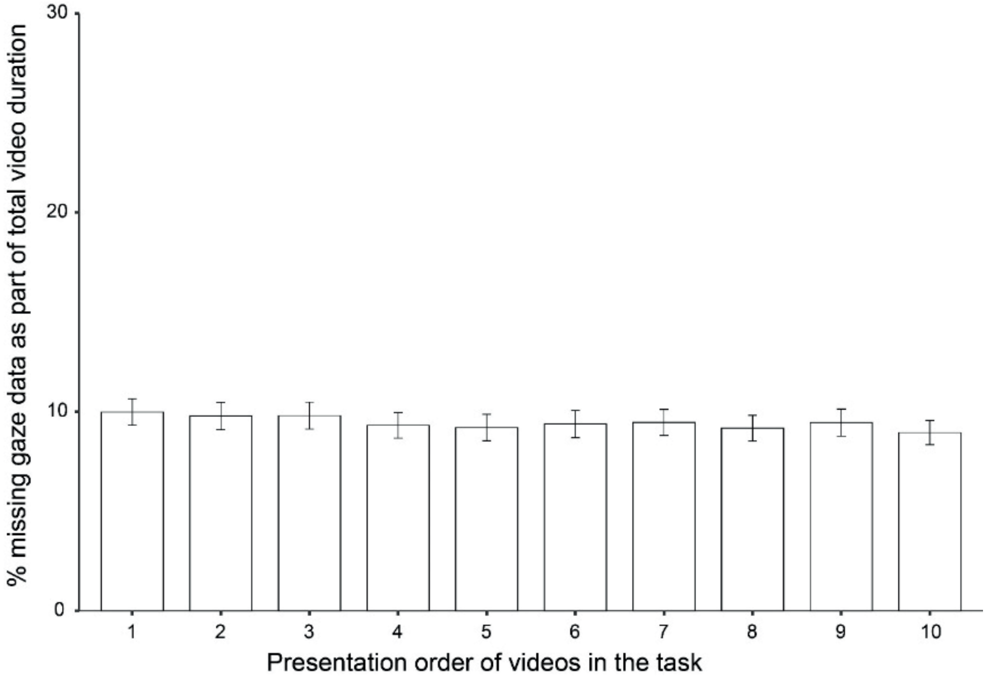
				Targets			
Videos	**Duration (s)**	**Valence**	***n***	**Target nr.**	**BEQ**	**Age**	**Gender**
1	116	N	98	1	3.5	25	Male
2	83	P	97	1	3.5	25	Male
3	107	N	95	2	5.03	26	Female
4	104	N	96	2	5.03	26	Female
5	155	N	101	3	4.06	62	Male
6	103	P	99	3	4.06	62	Male
7	102	P	99	4	5.97	24	Female
8	105	P	100	5	4.03	26	Male
9	121	N	96	6	3.61	23	Female
10	109	P	100	6	3.61	23	Female

Note. BEQ = Berkeley Expressivity Questionnaire; EA = Empathic accuracy; N = Negative, P = Positive, S = Seconds.

SUPPLEMENT S6.9

Average levels of the percentage missing gaze data as part of the total video duration per video order as presented in the task

The percentage missing gaze data did not differ between the videos, indicating that there was not more missing gaze data at the start of the task compared to later moments in the task. This indicates that the missing gaze data did not depend on the time point within the task. Error bars represent standard errors of the mean.



SUPPLEMENT S6.10

Generalized linear mixed regression model to assess the main effects and interaction of valence (positive versus negative) and target expressivity (BEQ average scores) on perceivers' level of EA (Fisher z transformed)

	B	SE	df	*t*	*p*
Model 1					
Intercept[1]	0.948	0.071	173.4	13.29	<.001
% dwell time eye region	-0.003	0.002	212.2	-1.46	.146
Model 2					
Intercept[1]	0.522	0.090	389.0	5.81	<.001
% dwell time eye region	0.003	0.002	460.4	1.41	.158
$Valence_{positive}$	0.761	0.108	888.5	7.07	<.001
% dwell time eye region × $Valence_{positive}$	-0.009	0.003	892.8	-3.33	<.001
Model 3					
Intercept[1]	1.209	0.309	937.7	3.92	<.001
% dwell time eye region	-0.009	0.008	924.8	-1.11	.269
BEQ	-0.062	0.072	895.3	-0.87	.385
% dwell time eye region × BEQ	0.002	0.002	895.6	0.81	.416
Model 4					
Intercept[1]	2.085	0.481	905.7	4.33	<.001
% dwell time eye region	-0.008	0.012	899.3	-0.70	.483
$Valence_{positive}$	-1.105	0.602	885.3	-1.84	.067
BEQ	-0.370	0.112	887.8	-3.31	<.001
% dwell time eye region × $Valence_{positive}$	-0.006	0.016	884.5	-0.37	.711
% dwell time eye region × BEQ	0.003	0.003	887.6	0.97	.331
$Valence_{positive \times}$ BEQ	0.441	0.141	885.6	3.13	.002
% dwell time eye region × $Valence_{positive}$ × BEQ	-0.001	0.004	885.1	-0.21	.833

[1] The intercept includes perceivers' level of EA (Fisher *z* transformed) during negative autobiographical target stories. BEQ = Berkeley Expressivity Questionnaire; EA = Empathic accuracy; SE = Standard error.