Predicting empathy from gaze dynamics
Natural Interaction & Affective Computing Project

Alessia Cecere

Abstract

This project explores a possible correlation between gaze dynamics and empathy. The analysis is conducted by adapting gaze feature extraction as described in D'Amelio et al., 2023 to predict a level of empathy for subjects in the EyeT4Empathy dataset (Lencastre et al., 2022). Sections 1 and 2 provide an insight into the psychological background of the problem at hand, while Section 3 describes utilized dataset elements. Section 4 explains the methods applied for gaze feature extraction and empathy prediction; obtained results are shown and discussed in Section 5.

1 Empathy

Empathy is one of the most important tools of social interaction. Although everyone has an understanding of its meaning – connected to the vicarious experience of others' emotions – there is no general agreement on its definition. The main problem in defining empathy is that it is not clear whether it should only include the recognition of a certain emotion, or its experience too (Bennett, 1995). A general distinction can therefore be made between *cognitive empathy* and *affective empathy* (Davis, 1980).

Cognitive empathy can be assimilated to *Theory Of Mind*, i.e. the capability of having a representation of other people's minds, of what they believe and what they want (Gärdenfors, 2006). According to Gärdenfors, this happens at different cognitive levels, which include having a *theory of attention* and a *theory of intentions*. Having a theory of attention means understanding what someone else is looking at by the direction of his/her gaze, along with the capability of seeing that he/she sees something. The theory of intentions, on the other hand, refers to the understanding of the goal behind others' actions, i.e the ability to uncover latent variables influencing their behaviour.

Affective empathy involves instead a rapid acknowledgement of others' emotions on the basis of facial expressions, gestures and voice prosody, which causes a direct emotional response in the individual (Reniers et al., 2011). This can be associated with Gärdenfors *theory of emotions* level, at which the subject understands that someone else is in pain and feels compassion; he/she feels the same as other because the other's sensory impressions provoke a corresponding emotions in him/her as well. This kind of perception has been associated with mirror neurons, a class of neurons first discovered in the monkey premotor cortex, that activate both when the monkey executes an action and when it observes a similar action performed by another individual (Rizzolatti and Craighero, 2004).

2 Gaze & Empathy

Gaze plays a huge role in interacting with the world and communicating: whenever individuals use vision, eyes are a channel to gather information and a signal that conveys information to others. Eye movements determine and change retinal input so that some things can be seen better than others or not at all (Schütz et al., 2011); this has an impact on one's perception of the world, which in turn tunes actions and beliefs. It is, therefore, no surprise that gaze influences empathetic abilities too. According to Leslie (1987), empathetic cues are acquired thanks to three components: an ID (Intentionality Detector) inferring goals and desires, an EDD (Eyes-Direction detector) that establishes a diadic interaction through gaze direction, and a third component (Shared Attention Mechanism or SAM) combining first two, allowing triadic interactions and intention reading from gaze. Later on Perrett and Emery (1994) corrected the model by proposing a general Direction of Attention Detector (DAD) module, exploiting a combination of eye and head movements.

Experiments have shown (Baron-Cohen et al.,

2001) that individuals use gaze direction to recognize mental states such as thinking, desire and refer. In the same experiments autistic children – who show an abnormal use of gaze (Kanner, 1943/1973) – failed to recognize those mental states from gaze direction. It was also shown that, under conditions in which the goal of an action is uncertain, the first place children (and also adults) look for information to disambiguate the goal is the eyes.

3 Dataset

To assess empathy on the basis of gaze dynamics, the EyeT4Empathy dataset (Lencastre et al., 2022) was utilized. The dataset includes scanpaths from 60 subjects aged between 20 and 40 years, the majority of whom were students. The number of trials performed by each subject varies from a minimum of 4 to a maximum of 49. Of the 60 participants, 30 were monitored while performing a task, while the other half was in a condition of free-viewing.

The task-oriented experiment consisted in asking participants to use their eye-movements to write sentences on a letter cardbord (right image in Figure 1), after an initial training on eye-tracking and the way non-verbal-movement-impaired people rely on it to communicate.

In the free-viewing experiment, on the other hand, participants were looking at images composed of random pixels (left and middle images in Figure 1) and asked to indentify objects or patterns: each stimulus was shown twice, during one-minute interval and alternating between trials.

A total of 4.8×10^6 points was recorded, using *Tobii Pro X3–120* eye-tracker and *Tobii Pro Lab Presenter Edition* software. The dataset already provided a classification in fixations and saccades, performed by the native eye-tracker algorithm, described in detail by the manufacturer in Olsen, 2012. In a first interation, the algorithm classifies fixations and saccades by using a velocity threshold of 30°/s. Following this, it combines nearby fixations by relabeling a group of points between two fixations as a single fixation if they are separated in both space and time by no more than 0.5° and 75 ms, respectively. Finally, the algorithm applies a minimum fixation time of 60 ms, and any fixation shorter than this duration is discarded. Any data points that do not meet the criteria for being a saccade or a fixation are categorized as *unclassified*.

In addition to saccade and fixation coordinates, the dataset includes many other data. Of particular interest, the subject's pupil diameter during both fixations and saccades was considered for this project. Pupil diameter has in fact been identified as a potential marker of empathy (Harrison et al., 2007, Cosme et al., 2021).

All subjects answered an extended version of the QCAE empathy questionnaire (Reniers et al., 2011, Bhurtel et al., 2021) before and after the intervention. The additional questions were specifically conceived to measure the level of empathy towards movement-impaired people, and were discarded for the more general purposes of this project. Furthermore, only the questionnaires taken before the experiments and the associated scores were considered, to try limiting external variables in free-viewing and especially in task-oriented group.

The test contains 35 items (of which 19 measure cognitive empathy and 12 affective empathy), to which participants could answer on a 4-point Likert-Scale with response options *strongly agree, slightly agree, slightly disagree* and *strongly disagree*; the general empathy score is obtained summing up the scores obtained for single questions (1 point for *strongly disagree* up to 4 points for *strongly agree*), and can therefore vary from 35 to 140. Cognitive and affective empathy scores were obtained in the same way, but summing the scores of the relative items. Table 1 contains statistics on empathy scores for subjects in the dataset.

4 Methods

4.1 Gaze feature extraction

Among the many possible descriptions of gaze dynamics, *foraging models* relate it to animal's behavior when looking for food. Foraging problems like what patch too choose, how to find it and when to leave it (Bartumeus and Catalan, 2009) are assimilated to foraging for information in the cognitive space, where exploration and exploitation are performed through saccades and fixations respectively. This interpretation of gaze dynamics is well-suited for the purposes of this project too: correlation between foraging habits of animals and their personalities has been found in terms of activity, boldness and social role in a group (Toscano et al., 2016).

In this project, gaze feature extraction was performed using the approach proposed in D'Amelio et al., 2023, briefly summarized in the remaining part of this section.

Each gaze position at time t in a trajectory can be represented by the location vector $x(t)$: as a

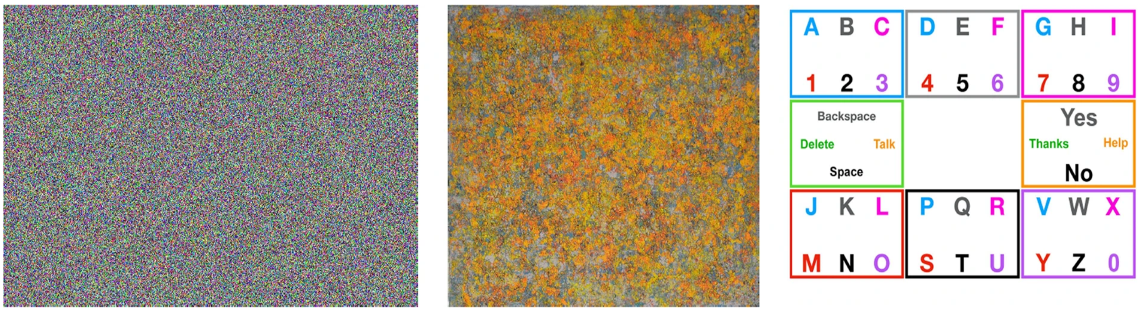

Figure 1: Stimuli shown to participants. Left and middle images were shown in the free-viewing experiment, the right one for task-oriented one.

Experiment	General Empathy	Cognitive empathy	Affective Empathy
Free viewing	95 ± 9.6	$58.6\ 4 \pm 7.5$	36.4 ± 4.3
Task oriented	94.9 ± 10.6	58.4 ± 7.7	36.5 ± 4.8

Table 1: Summary statistics (mean $\pm$ standard deviation) of empathy levels scored in experiment participants.

result, each observed trajectory is a realization of a stochastic process $X(t) = \{X(t), t = 0 \ldots T\}$, with $X(t) = x(t)$.

This trajectory is modeled using the Ornestein-Uhlenbeck (OU) process (Uhlenbeck and Ornstein, 1930), which appropriately describes the behaviour of a random particle, wandering but being pulled towards the current location of interest.

After the classification of events in fixations and saccades – which for this project was already provided by the Toby Eye-Tracker – the dynamical law governing the stochastic process evolution is set as

$$dx(t) = B^{S_t}(\mu^{S_t} - x(t))dt + \Gamma^{S_t} dW^{S_t}(t),$$

where

- $S_t \in [\text{sac}, \text{fix}]$, so that the prediction involves the next saccade if the current event is a fixation and vice-versa.
- $x(t)$ is the two-dimensional position in the trajectory, pulled towards the vector μ.
- B is a 2x2 matrix

  $$\begin{bmatrix} B_{ii} & B_{ij} \\ B_{ji} & B_{jj} \end{bmatrix},$$

  where B_{ii} and B_{jj} represent the drift towards the attractor and B_{ij} (which is equal to B_{ji}) the cross-correlation between the shift in the two dimensions, i.e. how much one is influenced by the other. The larger, the more curved the trajectory is.

Given an event e, the parameters $\theta = \{B^{st}, \Gamma^{St}\}$ are derived applying Automatic Differentiation Variational Inference (ADVI) to approximate the posterior probability

$$P(B^e, \Gamma^e | x^e) = \frac{P(x^e | B^e, \Gamma^e) P(B^e, \Gamma^e)}{P(x^e)}.$$

Resulting distributions are summarized by sample mean and highest density interval, and stacked together for Γ and B. For each subject's event, the following vector is thus obtained:

$$v_{id}^e = [B_{ii}^{\text{avg}}, B_{ii}^{\text{hdi}}, B_{ij}^{\text{avg}}, B_{ij}^{\text{hdi}}, B_{jj}^{\text{avg}}, B_{jj}^{\text{hdi}}, \Gamma_{ii}^{\text{avg}}, \Gamma_{ii}^{\text{hdi}}, \Gamma_{ij}^{\text{avg}}, \Gamma_{ij}^{\text{hdi}}, \Gamma_{jj}^{\text{avg}}, \Gamma_{jj}^{\text{hdi}}].$$

This vectorial representation was augmented by adding to both fixations and saccades the mean size of right and left pupil during the event; moreover, both saccade and fixation duration were inserted, along with saccade amplitude and angle. All features were normalized to be in a range between 0 and 1, to account for their different scales.

4.2 Empathy prediction

Once acquired a vectorial representation for each event, three Bayesian approaches were conceived, implemented and compared in order to predict a level of empathy from event features.

The first two models apply regression to predict a score on the QCAE test, while the third one classifies empathy in low and high. Distinct models were created for free viewing and task-oriented experiments, along with general, cognitive and affective empathy.

Subject events were split in train and test data at the level of the trial, with a ratio of 75% for training and 25% for testing. In all three cases, separate models were created for fixations and saccades, only later merged for the final prediction, i.e. the empathy value of a subject, given the events occurred in a trial.

Negative Binomial Regression

A first approach consisted in considering a Negative Binomial Distribution as underlying model generating observed empathy values. The likelihood for empathy values $E = \{e_1, e_2, \dots, e_n\}$ was thus assumed to be

$$P(e_i \mid \theta) = \mathrm{NegBin}(e_i \mid \lambda_i, \alpha)$$
$$\lambda_i = a + b^T x_i$$

where

- λ_i is the mean;
- α is the shape parameter of the Gamma distribution associated with the variance;
- x_i is the $F \times 1$ (F number of features) data point describing an event i (fixation or saccade).
- b is the $F \times 1$ vector of weights.

Both a and b have a normally distributed prior of mean 0 and standard variation 20, while α's prior is an Exponential of parameter 0.2.

Posterior probability was approximated with a Markov Chain Monte Carlo method, specifically the No-U-Turn Sampler (NUTS, Homan and Gelman, 2014). This sampling technique makes classical Metropolis-Hastings algorithm more efficient by exploiting Hamiltonian Monte Carlo simulation for sample proposal, but applying a recursive approach that reduces the need for expensive tuning.

To obtain a single predicted value for each test event, the mean of the values sampled from posterior distribution was selected. This process was repeated for all fixation and saccade events in the test data. Next, the event predictions were aggregated by selecting the mean predictions for saccades and fixations belonging to the same trial. Thus for each trial two mean values were obtained (one for saccades, one for fixations), finally averaged to arrive at a predicted empathy value for each (subject, trial) tuple.

Gaussian Mixture Regression

In contrast to the previous approach, the underlying model is here conceived as a mixture of two Gaussians.

$$\begin{aligned} P(e_i \mid \theta) = \pi_1 N(e_i \mid \mu_{(1,i)}, \sigma_1) \\ + \pi_2 N(e_i \mid \mu_{(2,i)}, \sigma_2) \end{aligned}$$
$$\mu_{(1,i)} = a_1 + b_1^T x_i$$
$$\mu_{(2,i)} = a_2 + b_2^T x_i$$

Both gaussian components have a uniform prior probability to contribute to the mixture, described by the weight vector $\pi = [\pi_1, \pi_2]$, having a Dirichlet distribution of parameters $[1, 1]$ as prior. Gaussian means are set as a linear combination of event features and associated weights.

Weight priors are normal distributions of mean 0 and standard deviation 10, while intercepts a_1 and a_2 have normal priors with standard deviation 5 and mean adapted to the distribution of observed empathy values, often organized in two visible clusters. Σ's priors – forming the vector $\sigma = [\sigma_1, \sigma_2]$ – are half-normals of standard deviation 10.

Logistic regression

In the third and last approach, the prediction problem was reduced to a binary classification in low and high empathy, with the application of logistic regression. The threshold was almost always set to mean empathy value; only in some cases it was identified as the median, in order to reduce the impact of outliers. After label assignment, the model was set as

$$P(e_i \mid \theta) = \mathrm{Bernoulli}(e_i \mid \sigma(l_i))$$
$$l_i = b^T x_i,$$

where σ is the logistic function and b is assigned a normal prior of mean 0 and standard deviation 10.

Final empathy prediction was obtained – as before – by taking the mean of posterior distribution samples for each event (fixation or saccade), computing the mean of events belonging to the same trial and taking the average of the two resulting fixation and saccade values, to get a score between 0 and 1 for each (subject, trial) tuple. This score was set to 0 (meaning low empathy) if lower than 0.5 and to 1 (high empathy) otherwise.

5 Results

Table 2 shows the accuracy obtained for classification in high and low empathy by means of logistic regression. *Fix* and *sac* refer to the accuracy of model predictions based on single fixation and saccade features, while *comb* indicates the accuracy after aggregation.

It can be noticed that combined predictions tend to have higher accuracy than event ones, which likely means that the aggregated model is learning different rules from different gaze dynamics. Furthermore, this shows that aggregating events is making subject classification more robust.

It can also be observed that saccade classification tends to have higher accuracy than fixation one. This can be explained by the fact that saccades have a better pointwise description, including the additional features of length and amplitude. Moreover, looking at Figure 2 it can be visualized that saccade amplitude has a higher – although always slight – correlation to empathy than other saccade features. Collaterally, higher correlation of pupil diameter can be visualized too: this property is consistent and present also in fixations, as expected based on literature.

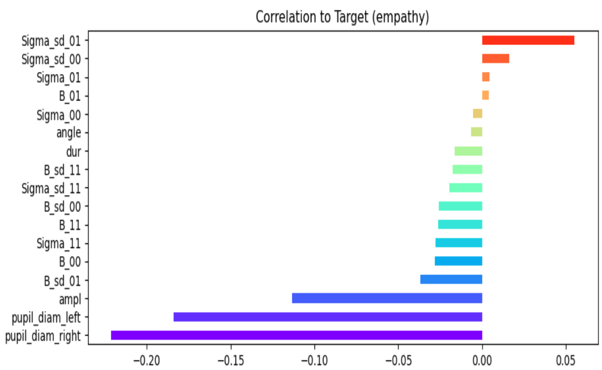

Figure 2: Visualization of correlation between different saccade features and general empathy in the free-viewing context.

Tables 3 and 4 show a comparison between results obtained with Negative Binomial Regression and Gaussian Mixture Regression, for free-viewing and task-oriented experiments respectively. To better interpret RMSE, one should have in mind that – even if the general empathy scale provided by the QCAE test has values between 35 and 140 – most of subjects in the dataset (and in original experiments) scored between 80 and 115. Cognitive and affective empathy scores tend to be even more uniform.

Taking a look at actual predictions, both Negative Binomial Regression and Gaussian Mixture ones tend to be around the mean, with a difference of maximum two points in worst cases: predictions almost or completely discard the already limited variety seen in the dataset. This phenomenon is also partially caused by the multiple applications of the average, which explains why – as opposed to classification results – combined predictions tend to be worse than event-based ones. Additionally, sampling problems emerged in both models. With Negative Binomial Regression, features were always assigned a low weight (almost zero) while in contrast α's values were higher than 500. This problem with feature weights was solved by Gaussian Mixture Regression, in which more reasonable weights were assigned. On the other hand, the sampling algorithm didn't manage to converge – as expected – in cases in which the two Gaussian components were not evident. As a result, Negative Binomial tends to perform better in frequency distributions that look like the one in Figure 3, while Gaussian Regression proves to be better with shapes like Figure 4, where more than one cluster emerge.

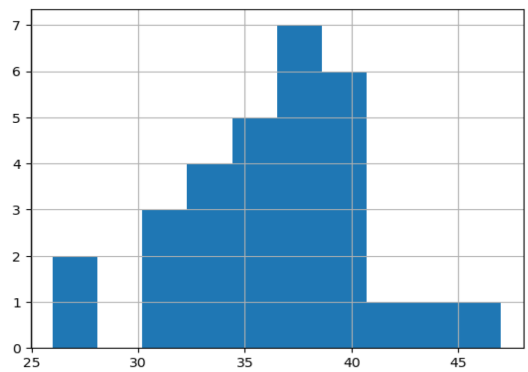

Figure 3: Frequency distribution of affective empathy for free-viewing experiment.

A general property that can be observed across all models is that results tend to be better for free-viewing than for task-oriented experiments. This may be due to the fact that during free-viewing experiments participants could exhibit a greater variability in eye movements, while when they were asked to perform a task – i.e. write on the whiteboard – their movements were more constrained. According to the cognitive relevance hypothesis, eyes are driven by top-down factors that intentionally direct fixations toward informative task-driven locations. In the absence of such task demands

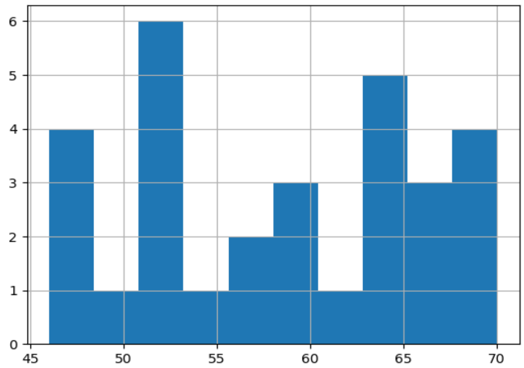

Figure 4: Frequency distribution of cognitive empathy for task-oriented experiment.

(as in scene-free-viewing) eyes are directed to low-level image discontinuities such as bright regions, edges, and color – the so-called salient regions (Borji and Itti, 2014). Tasks (like searching for a specific stimulus, in this case letters to express requested sentences) can suppress the influence of salience completely (Schütz et al., 2011). It is possible – although speculative – than in this greater variability a correlation with empathy emerged. It is also possible that the descriptive power of OU-Features was higher in free-viewing, which better matches foraging behaviour.

We can also globally observe that general empathy predictions tend to be better than the more specific ones, and that affective empathy is better predicted than cognitive empathy.

Conclusions

This project provided interesting insights into the complex problem of inferring empathy from gaze dynamics. Three Bayesian approaches were conceived, implemented and compared. Classification in high and low empathy proved to be more effective and robust – particularly for general empathy prediction – than regression methods trying to forecast an exact score on the QCAE test. Classification accuracy grew when aggregating different events (saccades and fixations) occurred during the same trial. Better performance was globally obtained analyzing free-viewing experiments compared to task-oriented ones.

Further research, also in the context of the master thesis, could be devolved to explore a correlation between what subjects are looking at in a context of free-viewing and their empathy scores. An other approach could be analyzing scanpaths during less constrained tasks – in which different gaze dynamics may reflect different solving strategies – and see if a correlation emerges in that case.

From the methodological point of view, more granular classification could be attempted, for example adding a *medium empathy* level; also other methods for regression could be examined.

Moreover, the same technique could be applied to infer different characteristics, like personality traits and affective states.

Experiment	General Empathy	Cognitive empathy	Affective Empathy
Free viewing	fix: 0.63 sac: 0.65 comb: 0.70	fix: 0.54 sac: 0.64 comb: 0.65	fix: 0.65 sac: 0.66 comb: 0.70
Task oriented	fix: 0.46 sac: 0.54 comb: 0.56	fix: 0.49 sac: 0.52 comb: 0.44	fix: 0.56 sac: 0.54 comb: 0.55

Table 2: Prediction accuracy using Logistic Regression; *fix* accuracy refers to empathy prediction from single fixation, *sac* from single saccade and *comb* refers to the accuracy of the final prediction for the (subject, trial) tuple.

Model	General Empathy		Cognitive Empathy		Affective Empathy	
	RMSE	MAPE	RMSE	MAPE	RMSE	MAPE
Neg Bin Regression	fix: 8.2 sac: 8 comb: 8.7	fix: 6.9% sac: 6.7 % comb: 7.2 %	fix: 7 sac: 7 comb: 7	fix: 10.2% sac: 10.1 % comb: 10.1	fix: 3.5 sac: 3.8 comb: 3.7	fix: 7.3% sac: 8.3% comb : 7.8%
Mix Gauss Regression	fix: 8.1* sac: 8* comb: 8.7*	fix: 6.9%* sac: 6.7%* comb: 7.3%*	fix: 7.3 sac: 7 comb: 7.6	fix: 10.7% sac: 10.2 % comb: 12.5	fix: 3.6* sac: 3.8* comb: 3.9*	fix: 7.5%* sac: 8.3%* comb: 8.7%

Table 3: Prediction RMSE and MAPE using Negative Binomial and Gaussian Mixture regression for **free viewing events**. Values with a * indicate that problems occurred during sampling, leading sampler not to converge.

Model	General Empathy		Cognitive Empathy		Affective Empathy	
	RMSE	MAPE	RMSE	MAPE	RMSE	MAPE
Neg Bin Regression	fix: 10.1 sac: 10.2 comb: 10.1	fix: 9.9% sac: 9.8 % comb: 9.9%	fix: 7.9 sac: 7.9 comb: 7.7	fix: 12.2% sac: 12.6 % comb: 12.7%	fix: 4.2 sac: 4 comb: 3.9	fix: 8% sac: 8.3% comb: 8.7%
Mix Gauss Regression	fix: 9.9 sac: 10.3 comb: 10	fix: 9.6% sac: 9.9% comb: 9.9%	fix: 7.6 sac: 7.9 comb: 7.6	fix: 12.2% sac: 12.5% comb: 12.5%	fix: 4.3 sac: 4 comb: 3.9	fix: 8.5% sac: 8.3% comb: 8.7%

Table 4: Prediction RMSE and MAPE using Negative Binomial and Gaussian Mixture regression for **task-oriented events**.

References

S Baron-Cohen, S Wheelwright, J Hill, Y Raste, and I Plumb. 2001. The "reading the mind in the eyes" test revised version: a study with normal adults, and adults with asperger syndrome or high-functioning autism. *J Child Psychol Psychiatry*, 42(2):241–251.

F Bartumeus and J Catalan. 2009. Optimal search behavior and classic foraging theory. *Journal of Physics A: Mathematical and Theoretical*, 42(43):434002.

J A Bennett. 1995. Methodological notes on empathy: further considerations. *ANS Adv Nurs Sci*, 18(1):36–50.

Samip Bhurtel, Pedro G. Lind, and Gustavo B. Moreno e Mello. 2021. For a new protocol to promote empathy towards users of communication technologies. In *Communications in Computer and Information Science*, pages 3–10. Springer International Publishing.

Ali Borji and Laurent Itti. 2014. Defending Yarbus: Eye movements reveal observers' task. *Journal of Vision*, 14(3):29–29.

Gonçalo Cosme, Pedro J Rosa, César F Lima, Vânia Tavares, Sophie Scott, Sinead Chen, Thomas DW Wilcockson, Trevor J Crawford, and Diana Prata. 2021. Pupil dilation reflects the authenticity of received nonverbal vocalizations. *Scientific Reports*, 11(1):1–14.

Mark Davis. 1980. A multidimensional approach to individual differences in empathy. *JSAS Catalog Sel. Doc. Psychol.*, 10.

Alessandro D'Amelio, Sabrina Patania, Sathya Bursic, Vittorio Cuculo, and Giuseppe Boccignone. 2023. Using gaze for behavioural biometrics. *Sensors*, 23(3).

Peter Gärdenfors. 2006. Reading other people's minds. In *How Homo Became Sapiens: On the evolution of thinking*. Oxford University Press.

Neil A Harrison, C Ellie Wilson, and Hugo D Critchley. 2007. Processing of observed pupil size modulates perception of sadness and predicts empathy. *Emotion*, 7(4):724.

Matthew D. Homan and Andrew Gelman. 2014. The no-u-turn sampler: Adaptively setting path lengths in hamiltonian monte carlo. *J. Mach. Learn. Res.*, 15(1):1593–1623.

Pedro Lencastre, Samip Bhurtel, Anis Yazidi, Gustavo B. M. e Mello, Sergiy Denysov, and Pedro G. Lind. 2022. Eyet4empathy: Dataset of foraging for visual information, gaze typing and empathy assessment. *Scientific Data*, 9(1):752.

Anneli Olsen. 2012. The tobii ivt fixation filter algorithm description.

Renate L E P Reniers, Rhiannon Corcoran, Richard Drake, Nick M Shryane, and Birgit A Völlm. 2011. The QCAE: a questionnaire of cognitive and affective empathy. *J Pers Assess*, 93(1):84–95.

Giacomo Rizzolatti and Laila Craighero. 2004. The mirror-neuron system. *Annu Rev Neurosci*, 27:169–192.

Alexander C. Schütz, Doris I. Braun, and Karl R. Gegenfurtner. 2011. Eye movements and perception: A selective review. *Journal of Vision*, 11(5):9–9.

Benjamin J Toscano, Natasha J Gownaris, Sarah M Heerhartz, and Cristián J Monaco. 2016. Personality, foraging behavior and specialization: integrating behavioral and food web ecology at the individual level. *Oecologia*, 182(1):55–69.

G. E. Uhlenbeck and L. S. Ornstein. 1930. On the theory of the brownian motion. *Phys. Rev.*, 36:823–841.