JUSTUS-LIEBIG-
UNIVERSITÄT
GIESSEN

Individual differences in gaze and neural representations

Synopsis zur kumulativen Dissertation
zur Erlangung des Doktorgrades der Naturwissenschaften
vorgelegt von Petra Borovska
am 13.03.2025

Justus-Liebig-Universität Gießen
Fachbereich 06
Psychologie und Sportwissenschaft

Erstbetreuer: Prof. Dr. Benjamin de Haas

Zweitbetreuer: Prof. Dr. Jutta Billino

Acknowledgments

I am deeply grateful to have been a part of this extraordinary environment for several years. It has shaped me in countless ways and motivated me to exercise my curiosity – the result of which is this thesis.

My special thanks belong to the Indivisual group, especially Ben, whose unwavering enthusiasm and unlimited support made everything possible. Your dedication and kindness fostered not only an ideal environment for scientific discussion and progress but also for friendliness and fun. Your patient encouragements motivated me to keep finding new ways, even where I could not see any, and I am incredibly grateful that you guided me through this PhD.

I am also immensely thankful for my exceptional colleagues, Max, Diana, and Marcel. Sharing an office with you meant sharing not just a workspace but the highs and lows of this journey. We discussed and troubleshooted, traveled, and presented, and grew together in our individual way yet always as a team. You are an amazing group, and I feel so fortunate to have been part of it.

I am also thankful to our many Hiwis, with special thanks to Diana Weissleder and Karin Schmidt, whose efforts to organize and ease our work are very much appreciated. I would also like to thank our collaborators, visiting researchers and interns, for sharing their expertise and valuable insights. My deepest gratitude belongs to my friends and family for their patience, support, and encouragement on my journey.

Finally, I would also like to thank the ever-growing psychology department for harboring people of various backgrounds and skills willing to share and discuss different topics day and night. This has been a truly inspiring environment.

Summary

This thesis addresses two fundamental yet distinct questions from complementary perspectives: one focuses on the average observer and the other on the individual. First, we asked whether the ability of faces to modulate gaze dynamics even pre-saccadically (i.e. before moving the eyes to a new gaze location) generalizes to natural scene viewing. Second, we investigated whether individual differences in gaze lead to individual neural representations of complex movie stimuli. Both studies are centered on examining visual behavior under naturalistic conditions.

In the first study, we investigate gaze dynamics in complex scenes, revealing that saccades toward faces are of higher velocity than those directed at inanimate objects, and fixations preceding face targets are shorter, especially when the face is near the current fixation point. These findings suggest that visual processing mechanisms prioritize faces even in naturalistic settings. This is especially remarkable given that gaze dynamics are in a complex scene exposed to objects embedded in visual clutter and the processing of the current and upcoming targets happens in parallel.

The second study examined how individual gaze behavior contributes to neural representations. We derived cross-brain decoding accuracy across pairs of observers using hyperalignment while participants viewed a movie under free-viewing and fixation conditions. The results showed that free-viewing led to increased neural activity compared to fixation. Moreover, the individual differences in gaze preferences, particularly fixation tendencies on faces and text, were related to variability in neural representations in the inferior temporal cortex (IT), and the individual differences in Euclidean gaze position contributed to neural divergence in both IT and the primary visual cortex (V1).

Together, these studies highlight the importance of high-level factors in gaze behavior and neural representations. We provide evidence that gaze rapidly orients toward semantically relevant features in a scene and that variability in preferential looking at specific semantic categories across observers can be linked to individual differences in the IT cortex. By examining gaze behavior across static and dynamic stimuli, this thesis underscores the role of active vision in shaping perception under naturalistic viewing conditions.

Contents

1 Introduction

In this thesis, two fundamental questions of vision science are investigated and supported by two studies. In the first study, we examine whether faces modulate gaze behavior during naturalistic viewing and whether this effect can be detected before a saccade is initiated. Crucially, this study addresses a problem of generalization from highly controlled paradigms with isolated stimuli to naturalistic free-viewing behavior and target behavior of a standard observer. In the second study, we diverge from a standard observer and focus on individual differences in gaze and neural representations. Here, we ask whether different individuals viewing identical dynamic movie stimuli have individually diverging neural representations in the inferior temporal cortex and beyond and whether this can be explained by individually diverging gaze. Although the two studies answer distinct questions, they are motivated by the effort to shift from highly controlled experiments to more naturalistic behavior.

1.1 Gaze dynamics in visual processing

The gaze behavior of a standard observer is highly complex. The visual system must adapt to a constantly changing visual environment while maintaining a stable representation of the world. Even when the external world remains relatively stable, extracting the necessary information to interpret this environment requires integrating multiple processes within the visual system. Rapid scene sampling is facilitated by eye movements, allowing for time-efficient processing of visual input. Due to differences in foveal and peripheral vision (Ludwig et al., 2014; Rosenholtz, 2016; Stewart et al., 2020) the gaze continuously moves the high-resolution fovea across the scene, shifting from one point of interest to another. This results in an alternating sequence of relatively stable fixations and rapid saccades, where speed plays a crucial role. For instance, during a group discussion, an observer continuously shifts their gaze between faces and gestures to pick up on important social cues, such as facial expressions, and use them to interpret and follow the conversation appropriately.

One advantage in this scenario is the visual system's ability to “pre-plan” a gaze shift by decoupling overt and covert attention (Grosbras et al., 2005; Talcott et al., 2023), enabling the processing of features of the upcoming target in parallel (Schwetlick et al., 2020). For instance, fixation duration is modulated by the low-level properties of an upcoming target, such as contrast and saturation (Einhäuser et al., 2020). Furthermore, a perisaccadic preview of the target is shifted along retinotopic coordinates during scene viewing, contributing to the decision of how long to remain in the current position before shifting to the next target (Schwetlick et al., 2020). Finally, transsaccadic predictions about an upcoming target

facilitate object recognition in the periphery for complex stimuli (Herwig & Schneider, 2014; Osterbrink & Herwig, 2021; Wilmott & Michel, 2021).

This mechanism also supports a larger task of the visual system: target selection. Which features dominate or are most useful for the fast processing of visual input? A substantial body of research has focused on low-level features of the scene, such as contrast and color, demonstrating that low-level saliency models can predict gaze behavior relatively well (Engmann et al., 2009; Harel et al., 2007; Itti & Koch, 2000). However, some predictors of gaze behavior are guided by eye movement tendencies that are relatively independent of the stimulus itself. These tendencies arise from the alternating sequence of eye movements; for instance, a saccade following a similar trajectory to the preceding one tends to be smaller (Tatler & Vincent, 2008), and fixation in between shorter (Nuthmann, 2017). Despite the importance of low-level features and oculomotor dynamics, they can only partially explain gaze behavior. As predictive models of gaze have advanced, it has become evident that low-level features alone are neither sufficient nor flexible enough to account for gaze dynamics fully (Kucharský et al., 2021; Kümmerer & Bethge, 2021; Tatler et al., 2017), particularly in more naturalistic scenarios (Roth et al., 2023; Tatler et al., 2011). Consequently, the focus has shifted to high-level and top-down factors that modulate gaze behavior.

1.2 Faces as gaze attractors

Object- and semantic-level information, in particular, dominates predictions of gaze behavior compared to those based on low-level aspects of a scene (Rubo & Gamer, 2018; Xu et al., 2014). Moreover, some semantic categories are stronger predictors than others. Specifically, faces and text have been shown to strongly modulate gaze across various contexts, from studies involving isolated stimuli (Boucart & Thorpe, 2016; Crouzet, 2010; Martin et al., 2018), to free-viewing of complex scenes (Cerf et al., 2009). It has been shown that faces are preferentially targeted (Coutrot & Guyader, 2014; Foulsham et al., 2010) and fixated on longer than other types of categories (Guo et al., 2006). The special status of faces is likely related to their behavioral relevance and the face diet we are exposed to from early development (Jayaraman & Smith, 2019). Face processing is managed by specifically dedicated brain areas, such as the fusiform face area (FFA) in the ventral cortex (Kanwisher, 2010). Moreover, this link is suggested to be a causal one (Parvizi et al., 2012). Although the concept of domain-specific face processing in the FFA might be an oversimplification (Vinken et al., 2023), it does not diminish the fact that faces play a crucial role in everyday life and are processed very efficiently by the visual system.

An even more striking example of efficient face processing is the modulation of saccadic velocity directed toward an upcoming face target. Contrary to the traditional view that the relationship between saccade velocity and amplitude is strictly stereotypical (Bahill et al., 1975), recent studies suggest that the relationship is rather modulatory and idiosyncratic (Reppert et al., 2015). In studies involving isolated faces, saccade velocities increased when directed toward face targets (Kauffmann et al., 2019), particularly in comparison to inanimate objects or random pixel noise (Xu-Wilson et al., 2009). Isolated faces can be regarded as high-value items (Yoon et al., 2020) carrying biologically relevant high-level information (Soares et al., 2017) that can trigger faster responses. However, it is unclear whether this effect generalizes in more complex situations where observers freely choose the target of the subsequent fixation.

In the first study, we utilized findings and mechanisms of gaze behavior in complex scenes to propose that faces are preferentially targeted, even in such environments, while controlling for many other factors (e.g., saccade amplitude). Specifically, we measured the velocity of saccades directed toward faces and the duration of fixation preceding saccades landing on faces and compared these metrics to saccades directed toward other objects.

However, it is increasingly evident that individuals are shaped by their visual world differently. Observers systematically differ in their gaze tendencies toward semantic categories in complex scenes (de Haas et al., 2019). For instance, while some tend to fixate more on faces, others preferentially target text. Moreover, the response time toward faces compared to other categories in the isolated stimuli experiment is consistently variable between observers, and this behavior can be linked to their gaze preferences towards faces in the free-viewing of complex scenes (Broda et al., 2024). These findings suggest that exploring similarities between observers is only one angle to approach this problem. Individual differences in gaze behavior provide an equally valuable perspective for understanding how visual information is processed.

1.3 Individual differences in dynamic stimuli

Most of the vision science research focuses on average and typical responses to visual stimuli of human observers, including our first study. The overarching aim is to describe and predict human behavior, assuming that we share more similarities than differences in our perception and that this perception objectively reflects reality. While this assumption holds to some extent, individual variability introduces two important considerations: first, the concept of an "average" observer is a simplification that neglects

systematic, and thus explainable, variance in the observed data, and second, the visual world may be more of a construct shaped by individual perception rather than a veridical reflection of reality.

In the second study, we thus shifted our focus to individual differences in neural representations and their relationship to variability in gaze behavior. This experiment not only changed the focus from average to individual differences but also examined variability in neural representations evoked by more complex dynamic stimuli, such as movies. Previous research has shown that individual differences in gaze behavior generalize from static to dynamic scenes (Broda & de Haas, 2022) and that Hollywood-style movies evoke a larger inter-individual coherence (Dorr et al., 2010; Hasson et al., 2010). This is in part due to the nature of movie stimulus. Naturalistic dynamic stimuli, like movies or spoken narratives, guide and direct gaze through carefully integrated scene cuts (Finn et al., 2020; Hasson et al., 2004, 2010; Nguyen et al., 2019). Part of this effect is achieved through rapid motion, which produces gaze convergence among viewers (Dorr et al., 2010; Hasson et al., 2010). The narrative also helps to sustain attention (Hasson et al., 2008) and directs focus to specific aspects of the movie, enhancing participant compliance (Vanderwal et al., 2019). In contrast, less engaging or incoherent movies result in more variable gaze patterns resembling the free-viewing of complex scenes (Dorr et al., 2010).

Movies, due to their unifying narrative structure, integrate a variety of complex visual and semantic information (Güçlü & van Gerven, 2017; Huth et al., 2012; Nishimoto et al., 2017). They offer a sufficient amount of data to characterize the neural patterns underlying visual perception (Feilong et al., 2018; Haxby, Guntupalli, et al., 2020), and other cognitive processes, such as memory (Furman et al., 2007; Jääskeläinen et al., 2021). Finally, using dynamic movie stimuli may improve predictions of human behavior compared to resting-state functional connectivity (Finn & Bandettini, 2021) and resolve feedforward from the feedback signal (Zhang et al., 2021). This makes movies and similar narrative stimuli particularly suitable for disentangling shared neural representations from those unique to individuals.

1.4 Aligning the brains

Variability between observers' neural response pattern can capture stable, trait-like characteristics (Geerligs et al., 2015) but can also reflect within-subject variability (Cutts et al., 2023; Hasson et al., 2010; Laumann et al., 2015), anatomical differences (Benson et al., 2022), or noise from data measurements. A relatively recent approach known as hyperalignment, first introduced by Haxby (2011; Haxby, Guntupalli, et al., 2020), addresses this variability by transforming individual voxel-wise

responses into a common representational space, aligning brains based on their shared functional structure rather than anatomical location. While the voxel-wise responses to the same stimuli differ across individuals, the representational geometry of these responses remains similar—especially when evoked by continuous stimuli, such as a movie, which provide rich, time-locked neural patterns that facilitate alignment.

Unlike traditional methods that average neural responses across voxels or map them into an abstract space using representational similarity analysis (Charest & Kriegeskorte, 2015; Kriegeskorte et al., 2008; Kriegeskorte & Kievit, 2013), hyperalignment captures finer-scale topographic features, providing a more precise and individualized representation of neural activity. Additionally, categories like animacy and faces are still represented in the shared representational space and overlap with functional representations derived from anatomical analyses (Feilong et al., 2018; Guntupalli et al., 2017; Jiahui et al., 2020). However, these categories play a relatively minor role in explaining variance in neural responses to rich, dynamic stimuli such as movies (Sha & Haxby, 2015). When hyperalignment is applied with MVPA on still images with a limited number of categories, the resulting shared representational space does not generalize well to movie stimuli. In contrast, a common space derived from neural activity during movie viewing can classify neural patterns evoked by categories from still images (Haxby et al., 2011; Sha & Haxby, 2015). This highlights the advantages of using dynamic stimuli and underscores the challenges of generalizing findings from controlled laboratory settings to real-world scenarios (Nastase et al., 2020).

The core of hyperalignment lies in the Procrustes method, which uses rotation and reflection to minimize the Euclidean distance between coordinates, thereby achieving cross-brain alignment. The outcome of this calculation is a transformation matrix, described by (Guntupalli et al., 2016) as a ‘key that unlocks that person’s neural code’. Through an iterative process of pairwise alignment, a common model for the entire group of subjects is constructed. This common model is a high-dimensional space that captures shared information while retaining the distributed fine-scale topography, allowing for a more sensitive uncovering of individual differences. Although hyperalignment reduces topographical variability, it makes individual differences in representational geometry more reliable across independent data samples than anatomical alignment (Haxby, Guntupalli, et al., 2020). This is because the transformation preserves information derived from the relationships among pattern vectors, ensuring that functionally equivalent neural responses are aligned across subjects, even if their original voxel locations differ

(Feilong et al., 2018). As a result, hyperalignment is well-suited for identifying individual differences and serves as a more stable predictor of behavior (Feilong et al., 2021; Jiahui et al., 2023). However, despite the reliability of fine-scale individual differences, they are less heritable compared to coarse-scale representations (Busch et al., 2024). This suggests that environmental factors influence idiosyncrasies in fine-scale topographies more than purely genetic ones (Busch et al., 2024).

The adaptation of the hyperalignment algorithm in a recent study (Jiahui et al., 2023) suggests that high accuracy can be achieved by simplifying the process. Rather than aligning all subjects to a common representational space, this method directly transforms one subject's representational space to match another's. Inspired by this approach, we similarly modified our hyperalignment algorithm, omitting the step of iteratively projecting each subject's representational space into a common space. Instead, we directly compared subjects in a pairwise manner and estimated representational divergence between pairs of observers based on cross-brain decoding accuracy (Borovska & de Haas, 2024). This resulted in an estimate of individual neural divergence between observers and allowed us to model this as a function of gaze divergence.

1.5 Individual gaze modulations of neural representations

It is not entirely clear what the effect of individual gaze on neural divergence is. Eye movements carry meaningful information reflecting idiosyncratic oculomotor signatures (Andrews & Coppola, 1999; Bargary et al., 2017). These signatures may represent stable traits (Mollon et al., 2017) and provide valuable insights into how gaze behavior interacts with neural representations of the visual world. Individual gaze shaped by anatomical and functional differences across early and higher stages of visual processing may contribute to the functional specificity of each individual's visual system. For instance, in a recent study by Arcaro et al. (2017), monkeys not exposed to faces during development failed to develop face patches in the inferior temporal (IT) cortex. This suggests that variability in visual experiences can have significant consequences, with individual gaze idiosyncrasies playing at least a partial role. Previous research (de Haas et al., 2019) shows that individuals' preferences for fixating on distinct semantic categories differ. These differences are stable, highly consistent, and cannot be explained solely by visual field biases. Notably, these preferred semantic categories align with domain-specific patches in the ventral stream (Grill-Spector & Weiner, 2014; Margalit et al., 2023), suggesting that the individual visual responses in the ventral cortex may be linked to individual gaze preferences. The well-established foveal bias of IT further supports this idea (Hasson et al., 2002; Kravitz et al., 2013;

Levy et al., 2001) with a tight link between early foveal visual areas and category-selective regions in the ventral cortex (Finzi et al., 2021; Himmelberg et al., 2022). Eye movements also facilitate higher cognitive functions, such as memory. For example, larger images are remembered better, an effect likely modulated by gaze behavior, as larger images tend to produce more saccades (Fehlmann et al., 2020; Liu et al., 2020; Masarwa et al., 2022). Factors such as saccade frequency, the coherence of scan patterns between encoding and recall, and the extent of visual exploration influence whether an image is remembered (Broers et al., 2022; Damiano & Walther, 2019; Ryan & Shen, 2020).

On the other hand, eye movements can often be treated in fMRI as a source of noise or included in the analysis as nuisance regressors (Lu et al., 2016). Research designs frequently account for this by instructing participants to fixate during stimulus presentation (Huth et al., 2012; Kay et al., 2008). However, recent studies suggest that this may not always be necessary. For instance, regions like the ventral cortex exhibit relative invariance to eye movements, minimizing the impact of constant gaze shifts (Nishimoto et al., 2017). Additionally, the IT cortex may reflect broader visual field coverage than previously assumed (Park et al., 2023), which might suggest a diminished role of eye movements on neural divergence. Moreover, research employing hyperalignment techniques during free-viewing of movie stimuli has demonstrated exceptionally high accuracy in between-subject multivariate pattern analysis within the ventral cortex (Feilong et al., 2022; Haxby et al., 2011; Jiahui et al., 2023; Visconti di Oleggio Castello et al., 2020).

Together, these findings suggest that the effect of individual gaze on neural representation is somewhat ambiguous and needs further clarification. The success of hyperalignment and the large visual field coverage of IT suggest that individual gaze might not matter much for visual response in the ventral cortex. On the other hand, the foveal bias in IT and the preference of individual gaze towards several distinct semantic categories that match the functional organization of IT might indicate that gaze may result in individual diverging ventral representations of naturalistic content. In the second study, we test whether it is true that different individuals watching the same movie may have divergent neural representations that could be explained by gaze idiosyncrasy.

2 Study 1: Faces in scenes attract rapid saccades

Borovska, P., & de Haas, B. (2023). Faces in scenes attract rapid saccades. Journal of Vision, 23(8), 1–15. https://doi.org/10.1167/jov.23.8.11

In this project, we investigated how free-viewing dynamics vary as a function of an upcoming fixation target while controlling for various low-level factors. We compared saccade velocity and duration between saccades directed toward faces versus inanimate objects.

We analyzed data from 101 participants free-viewing 700 complex everyday scenes. This resulted in thousands of valid event series of gaze that landed on a face or an inanimate object. We measured the saccade velocity toward face targets and the fixation duration of the current fixation before landing on a face. Given the complexity of gaze behavior during free-view, we used linear mixed-effects models to measure both saccade velocity and preceding fixation duration while controlling for several oculomotor and low-level factors. Based on previous research, we included the following predictors: (1) target saccade amplitude (Fig. 1), (2) incoming saccade amplitude (Fig. 1), (3) size of target stimuli, (4) time from onset of the trial, (5) angle of the target to incoming amplitude (Fig. 1), (6) GBVS at the intermediate fixation, and (7) GBVS at the target fixation.

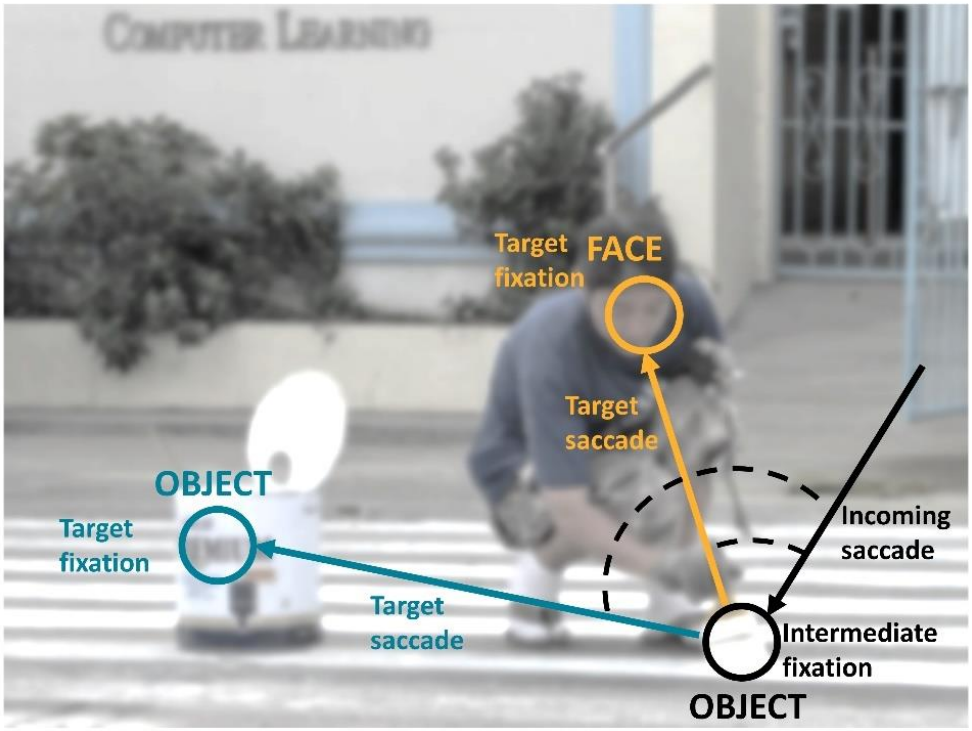

Figure 1. A sequence of incoming saccade, intermediate fixation, target saccade, and target fixation overlaid on an example image. We identified face-related and inanimate object–related saccades as target saccades that landed either on an inner face region (orange example) or an inanimate object (cyan example). Our dependent variables were the peak velocity of the target saccade and the duration of the preceding intermediate fixation. Independent control variables included the amplitude and peak velocity of the incoming saccade and the angle between the incoming saccade and the target saccade (dashed lines). Note the example image (cf. Xu et al., 2014) shown has been blurred for illustrative purposes.

By contrasting fixations landing on faces versus neutral objects, we confirmed that saccade velocity towards faces was higher as compared to other objects (Fig. 2 d) and the duration of the preceding fixation is shorter when faces are the upcoming targets (Fig. 3 d), mainly if the face target is close (Fig. 3 b). This is consistent with notions from transsaccadic literature (Osterbrink & Herwig, 2021; Wilmott & Michel, 2021), suggesting that features of upcoming targets have already been processed during the current fixation. Furthermore, recent evidence for a brief retinotopic shift of attentional window along the saccade trajectory (Schwetlick et al., 2020) seems to be more pronounced when the attentional window falls on a face in the parafovea (considered to span ~4-5° of eccentricity), as indicated by our results (Fig. 3 b).

These findings suggest that the dynamics of gaze behavior during free-viewing of complex scenes are modulated by several interacting factors which should be considered in the study of natural vision. Crucially, two separate findings from previous experiments with isolated stimuli, namely that of higher saccade velocity and lower latency for face targets, generalize to free-viewing of complex naturalistic stimuli.

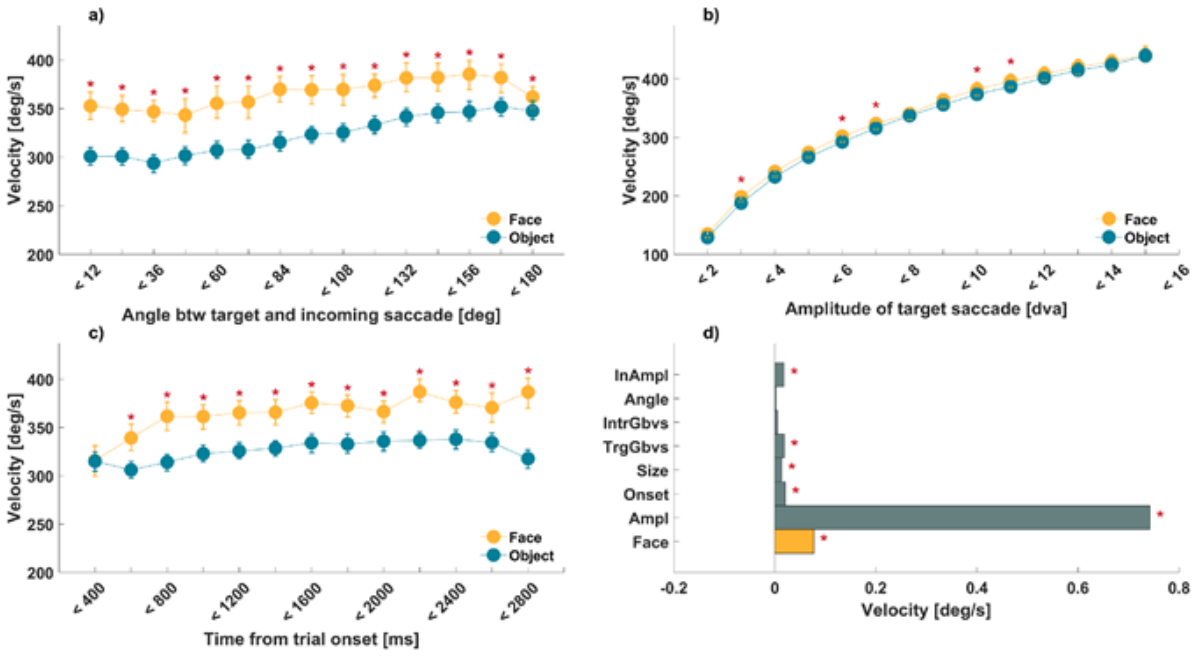

Figure 2. Peak velocity. (a–c) Peak velocity of target saccades landing on faces and inanimate objects in cyan and yellow, as shown in the inset. Red asterisks mark Bonferroni-corrected significance of paired t-test and error bars represent bootstrapped 95% confidence interval (1,000 resamples). (a) Peak velocity as a function of absolute deviation of saccade angles between target and incoming saccade. The increase in angle represents an increase from same-directed saccades (<12 degrees) to opposite-directed saccades (<180 degrees). (b) Peak

velocity as a function of target saccade amplitude (in dva), showing the main sequence. (c) Peak velocity as a function of time from onset (ms) within a trial. (d) Standardized, fitted predictor weights of a linear mixed-effects model of peak velocity with simple main effects of semantic target category (Face; shown in yellow bar), target amplitude (Ampl), time from trial onset (Onset), size of target stimuli (Size), low-level salience at target (TrgGbvs) and intermediate (IntrGbvs) fixation, absolute deviation of saccade angle between incoming and target saccade (Angle), and amplitude (InAmpl) of the incoming saccade. Red asterisks mark statistically significant beta coefficients. Note that all continuous variables were z-scored and thus the corresponding beta values indicate effects in standard deviation units.

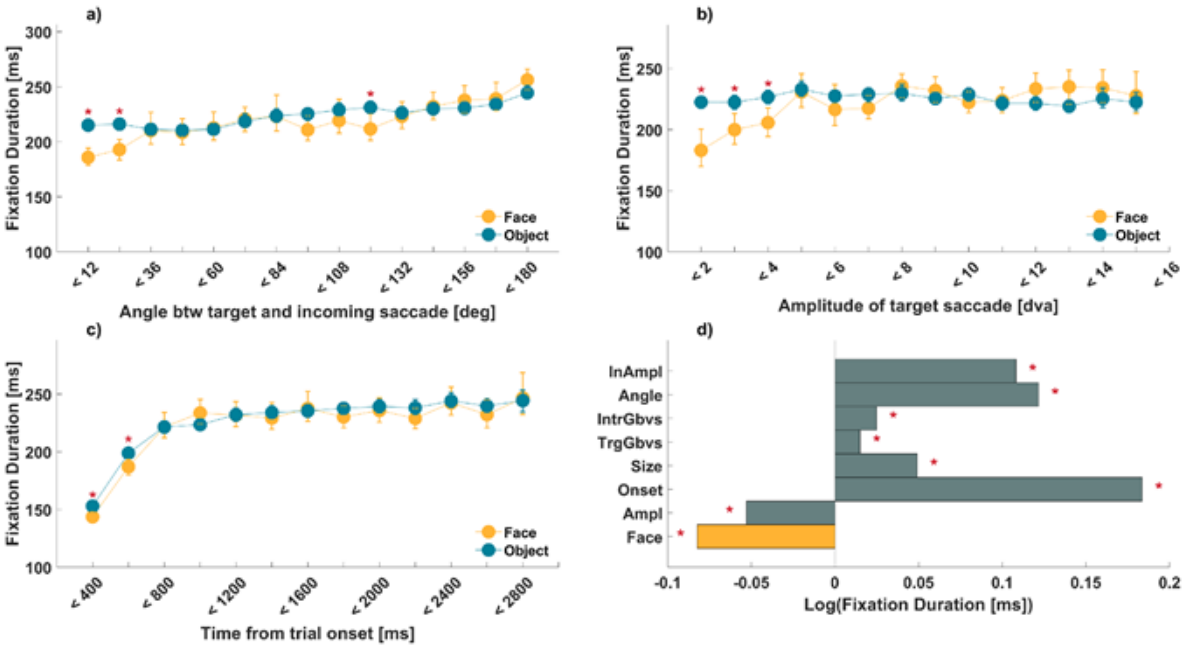

Figure 3. Measures of fixation duration. (a–c) Intermediate fixation duration (ms) followed by saccade landing on faces and inanimate objects in yellow and cyan, as shown in the inset. Red asterisks mark Bonferroni-corrected significance of paired t-test and error bars represent bootstrapped 95% confidence interval (1,000 resamples). (a) Intermediate fixation duration as a function of absolute deviation of saccade angles between target and incoming saccade. The increase in angle represents an increase from same-directed saccades (<12 degrees) to opposite-directed saccades (<180 degrees). (b) Intermediate fixation duration as a function of target saccade amplitude (in dva). (c) Intermediate fixation duration as a function of time from onset (ms) within a trial. (d) Standardized, fitted predictor weights of a linear mixed-effects model of intermediate fixation duration with simple main effects of semantic target category (Face; shown in yellow bar), target amplitude (Ampl), time from trial onset (Onset), size of target stimuli (Size), low-level salience of target (TrgGbvs) and intermediate (IntrGbvs) fixation, absolute deviation of saccade angle between incoming and target saccade (Angle), and amplitude of the incoming saccade (InAmpl). Red asterisks mark statistically significant beta coefficients. Note that all continuous variables were z-scored and thus the corresponding beta values indicate effects in standard deviation units.

3 Study 2: Individual gaze shapes diverging neural representations

Borovska, P., & de Haas, B. (2024). Individual gaze shapes diverging neural representations. *Proceedings of the National Academy of Sciences, 121*(36), 2017. https://doi.org/10.1073/pnas.2405602121

In this study, we tested the hypothesis that individually diverging neural representations can be explained by idiosyncrasies in gaze upon viewing identical complex dynamic stimuli. In particular, we focused on differences in average Euclidean distance between gaze positions, the tendency to fixate on face and text, and the differences in saccade amplitude and rate.

We used a machine learning technique, hyperalignment (Haxby et al., 2011; Haxby, Guntupalli, et al., 2020), to test this prediction, and derived a cross-brain accuracy used as a proxy for representational similarity between pairs of observers. In our experiment, we let participants watch a movie in two sessions, once with an eye-tracker and once in an fMRI scanner. The order of sessions was counterbalanced. During the scanning session, participants either fixated on the center of the screen while the movie was played or were allowed to free-view. We mainly focused on the inferior temporal cortex (IT) and primary visual cortex (V1) but also extended our analysis to the whole brain.

We could decode neural representations of complex visual stimuli across brains using hyperalignment. We found that free-viewing (compared to central fixation) leads to a substantial increase in BOLD signal amplitude across the visual system (Fig. 4 D). Crucially, we found that individual eye movements enhance cortical visual responses but lead to representational divergence in IT and V1 (Fig. 4 A). Moreover, pairwise differences in gaze explained pairwise differences in neural representations. Specifically, the average Euclidean distance between gaze positions predicted the representational divergence both in IT and V1 (Fig. 4 B), and the tendency to fixate on faces and text predicted neural divergence in IT but not in V1 (Fig. 4 B). Among predictors that did not predict the representational divergence in either IT or V1 were pairwise differences in saccadic amplitude and rate (Fig. 4 B). Finally, we expanded our analysis to the whole brain and found that the effects of neural divergence were mainly limited to the occipital and inferior temporal regions (Fig. 4 C).

These findings indicate that beyond idiosyncrasies of functional layout, which can be overcome by hyperalignment, individual differences in gaze lead to individual differences in ventral representations. Specifically, pairwise differences in the spatial distribution of gaze and semantic salience contribute to neural divergence in IT.

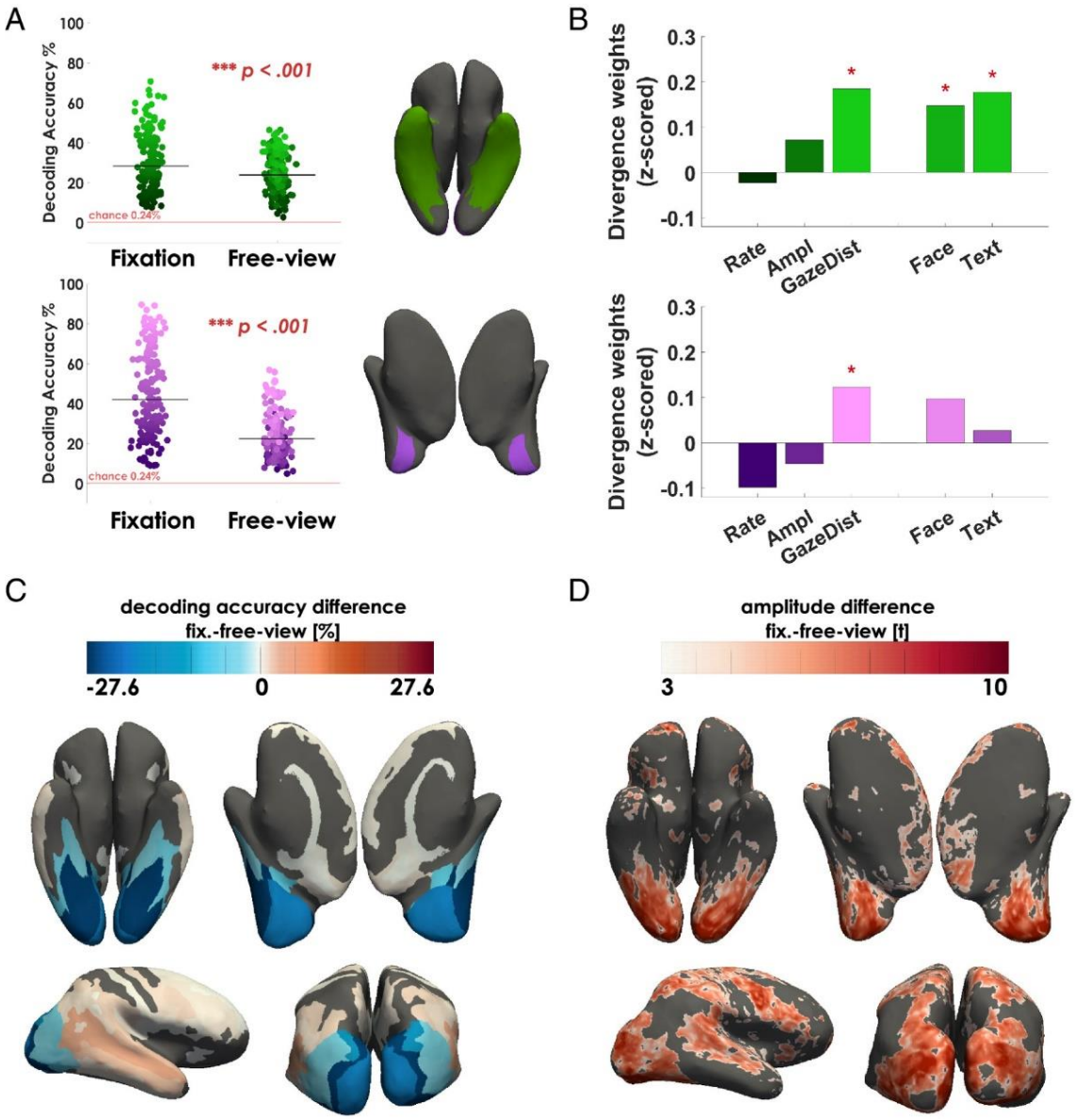

Figure 4. Results. (A) Cross-brain decoding accuracy in the fixation and free-viewing conditions for IT and V1. Each dot represents one pair of observers (N = 171 pairs), the chance level of 0.24 % is indicated by red lines, and P-values correspond to a GLMEs testing the effects of the conditions on cross-brain decoding accuracy. The Top (green) and Bottom (purple) plots show data from IT and V1 and the corresponding ROIs. The shades of dots correspond to decoding accuracy in the fixation condition. (B) Fitted weights of individual differences in low- and high-level gaze parameters predicting pairwise neural divergence in IT and V1. Simple main effects are shown for IT in green (Top) and V1 in purple (Bottom). Rate and Ampl: individual difference in the saccadic rate and amplitude; GazeDist: average Euclidean distance between gaze positions of two observers; Face and Text: individual difference in the tendency to fixate faces and text, respectively. Asterisks indicate significant simple main effects that survived Bonferroni correction. (C) Effects of free-viewing on neural alignment between observers. The heatmap shows the average difference in cross-brain decoding accuracy (in %) between free-viewing and fixation conditions across all pairs of observers on the inflated cortical surface. The color-coding indicates the magnitude of the difference for each region of interest (ROI), extracted using the Destrieux atlas parcellation and the Benson retinotopy atlas for V1, V2, and V3. Only differences that were significant at P < 0.001 according to the GLME are shown (SI Appendix, Supporting Methods). Values are color-coded as shown in the inset bar, with cool colors indicating a drop of decoding accuracy in the free-viewing condition. (D) Amplitude effects of free-viewing. The heatmap shows vertex-wise t-values for the contrast between free-viewing and fixation on the inflated cortical surface of an example observer. t values are color-coded as shown in the inset bar. Inflated hemispheres are shown in inferior (Top Left), medial (Top Right), lateral (Bottom Left), and posterior views (Bottom Right).

4 Discussion

In this section, we will discuss the results of two distinct approaches. The first study addresses a rather fundamental question of gaze modulations by face targets embedded in complex scenes. Our second study answers another essential question of visual neuroscience, whether individually diverging neural representations can be explained by gaze idiosyncrasy. We discuss these separately but point to their connections in the context of naturalistic paradigms at the end of this section.

4.1 Processing of high-level features during natural vision

Many studies investigating human behavior and neural responses using simplistic artificial stimuli assume that complexity is merely the sum of its parts. As a result, these findings are often expected to generalize to more complex, real-world situations. Consequently, one line of reasoning suggests that naturalistic stimuli should be tested post hoc based on results from highly controlled and constrained experimental designs (Rust & Movshon, 2005). Even if this were a fully adequate assumption, the links between impoverished artificial stimuli, complex natural stimuli, and real-world behavior still need to be empirically tested - particularly in cases where findings do not perfectly align between simple and complex settings.

Here, we tested whether the effect of saccades with lower latency and higher velocity shown in studies using isolated stimuli (Broda et al., 2023; Crouzet, 2010; Xu-Wilson et al., 2009), generalizes to naturalistic human behavior while viewing a complex scene. Indeed, we found that even during free-viewing, where the gaze is exposed to complex stimuli marked by visual clutter and concurrent processing of current and upcoming targets, saccades toward faces are faster, and the duration of preceding fixation is shorter. These effects were significant while considering a range of low-level factors shown by previous research to modulate gaze dynamics.

Although rapid saccades toward faces generalize from isolated stimulus experiments to free viewing, as demonstrated in our first study, the processing time is slower during free viewing (200 ms for fixations smaller than 4° of eccentricity; (Borovska & de Haas, 2023)) compared to 140 ms in isolated conditions (Crouzet, 2010). This discrepancy may arise partly due to the complexity of natural scenes, where multiple potential targets and distractors are simultaneously present. Multiple competing stimuli increase the need for a pre-saccadic shift of attention, which consequently prolongs fixation duration—a phenomenon that may be diminished in simpler settings (Talcott et al., 2023). This could also be

explained in part by specific instructions and the choice of a particular task. In studies with a two-alternative forced-choice paradigm (2-AFC), participants are instructed to saccade as fast as possible, which may additionally decrease the latency (Crouzet, 2010).

We demonstrated an effect of upcoming semantic information on saccade velocity and the fixation duration of the previous fixation. Unlike studies using isolated stimuli, where a speed advantage for faces persists even at extreme eccentricities of 80° (Boucart & Thorpe, 2016), the effect on fixation duration during free-view of complex scenes was limited to faces in the parafovea (within 4° of eccentricity). In contrast, the effect on saccade velocity extended farther into the periphery (up to 11° of eccentricity). We hypothesized that fixation duration and saccade velocity rely on separate mechanisms during free-viewing of complex scenes. This could be partially driven by visual field coverage that may be broader for saccade velocity and possibly linked to the subcortical pathway via superior colliculus (Beltramo & Scanziani, 2019; Nguyen et al., 2014; Soares et al., 2017). In contrast, fixation duration is likely associated with the central bias observed in occipitotemporal processing (Hasson et al., 2002), which suggests preferential processing of faces in the fovea. Findings from a study with monkeys further support this (Krishna et al., 2014) by showing enhanced foveal processing when the face target is closer to the current fixation. This might suggest that providing a larger face at farther eccentricities might yield similar effects as with faces occurring in the parafovea. However, the interaction of eccentricity and size of the face stimulus should be further explored, especially in the context of more naturalistic scenarios in which the effects of crowding are stronger (Rosenholtz, 2016) and the visual system is more reliant on predictions about upcoming foveal input (Hayhoe et al., 2012; Huber-Huber et al., 2021).

4.2 Diverging neural representations and their link to gaze behavior

As shown earlier and discussed further in this section, individual variability in gaze and brain is a crucial factor to consider. While neural activity tends to converge among observers in visual regions (Hasson et al., 2004; Laumann et al., 2015) and gaze may not matter much for neural representations in IT (Nishimoto et al., 2017; Park et al., 2023), our findings clearly demonstrate that variability in neural representations in both low- and higher-order visual areas is at least partially driven by differences in gaze. We show that while the amplitude of the BOLD signal increases during the free-view condition, the cross-brain decoding accuracy decreases in the free-view condition compared to the condition where participants are instructed to fixate, and this difference is likely driven by gaze divergence. Specifically, we found that the pairwise Euclidean distance of gaze positions between observers predicted

representational divergence in both V1 and IT, whereas pairwise differences in the tendency to fixate faces and text predicted the neural divergence in IT only.

The increase in the amplitude of the BOLD signal during the free-viewing compared to the fixation condition aligns with findings of increased neural activity evoked by gaze shifts (Lu et al., 2016; Parker et al., 2023; Xiao et al., 2024). Although the amplitude of the BOLD signal was higher in IT during the free-viewing condition, cross-brain decoding accuracy was lower for free-viewing compared to the fixation condition. This suggests that eye movements likely drive differences between conditions and cannot be explained solely by the strength of the BOLD signal.

Previous studies indicate that early visual areas are highly retinotopic (Benson et al., 2012; Dougherty et al., 2003) with foveal input being overrepresented due to cortical magnification (Qiu et al., 2006). This indicates that gaze shifts induced by eye movements have a significant impact on activity in these areas (Nishimoto et al., 2017; Parker et al., 2023). However, it remains unclear how sensitive higher-level visual areas, such as IT, are to eye movements. The IT cortex exhibits some degree of retinotopic organization (Groen et al., 2017; Silson et al., 2015; Xiao et al., 2024), for instance, stronger activation has been observed for contralateral compared to ipsilateral visual field stimuli (Silson et al., 2022) and position information can still be extracted from category-selective regions in IT (Schwarzlose et al., 2008). Additionally, the inactivation of face-selective neurons in IT directly affects eye movements and alters scan patterns on faces (Azadi et al., 2024). On the other hand, IT appears to be less sensitive to eye movements (Nishimoto et al., 2017) and has been implicated in the stabilization of visual input (Piasini et al., 2021). If neural representations in IT were fully invariant to eye movements, we would expect little to no difference between conditions in our study and no link between gaze behavior and IT representations. While the difference between our two conditions (central fixation vs. free-view) was smaller in IT than V1, it was significant in both areas. However, only semantic salience and not our predictors for low-level oculomotor gaze parameters, i.e., differences in saccadic amplitude and rate, significantly contributed to neural divergence in IT. While Nishimoto et al. (2017) found IT relatively insensitive to eye movements, our results suggest that IT, albeit less sensitive to eye movements than V1, remains significantly affected with responses driven more by semantics than oculomotor dynamics.

To link differences in gaze to differences in neural representations, we built two models for each region of interest (ROI), IT and V1, incorporating several predictors that quantify pairwise differences in gaze. The strongest predictor in our models for both V1 and IT was the average Euclidean distance between

gaze positions across observers. Previous research on gaze behavior during free viewing of dynamic stimuli suggests that motion content (Russ & Leopold, 2015; Smith & Mital, 2013), biological motion (Haxby, Gobbini, et al., 2020) and scene cuts (Hasson et al., 2008) might be key factors influencing fixation, likely reducing interindividual variability (Dorr et al., 2009). Including these features in our models as additional predictors could improve the model's explanatory power, particularly concerning differences in the average Euclidean distance between gaze positions of observers.

Based on previous research on individual semantic biases in gaze (de Haas et al., 2019) and the category selectivity of IT (Bracci et al., 2017; Grill-Spector & Weiner, 2014), we included individual face and text gaze preferences as predictors of neural divergence in IT. We could predict neural divergence based on face and text gaze saliency in a separate model, though with a smaller effect than differences in Euclidean gaze position. This finding aligns with the presence of foveal face- and text-preferring neurons in IT (Grill-Spector et al., 2017; Hasson et al., 2002; Silson et al., 2022), but also with evidence that face responses account for only a small portion of the variance in neural activity elicited by complex movie stimuli (Haxby et al., 2011). This study aimed to establish the relationship between individual gaze patterns and individual neural representations. As an initial step, this objective can be considered achieved.

Our results show that the unique way we look at stimuli matters for their neural representations. Previous studies suggest that individual gaze may even be optimal for our own processing. For example, in a study by Peterson et al. (2013), they show that gaze diverging from the canonical point of fixation does not lead to detrimental effects on performance in a face-identification task. Additionally, an individual's gaze pattern on faces and houses better predicts their neural activity in IT compared to other subjects' brains (Wang et al., 2019). On the other hand, in some cases, there might be an advantage in becoming more similar to someone else. For instance, in a study by Meshulam et al. (2021), they found that aligning one's neural representation more closely with that of experts enhances students' learning outcomes, and a larger gaze dispersion on an image has been linked to better memorability of a scene (Broers et al., 2022). This suggests that certain aspects of shared gaze behavior may be beneficial, while some idiosyncrasies could be somewhat maladaptive. To dissociate the advantages of canonical versus idiosyncratic gaze behavior more clearly would be rather intriguing. .

4.3 Limitations

Our studies have several limitations that could be addressed or extended in future research. In our first study, we calculated saccade velocity and preceding fixation duration towards faces and averaged these measures across observers. Previous research suggests large individual variability in gaze behavior (de Haas et al., 2019) and recent findings show that the individual differences in latency toward faces correlate to face salience during free-viewing of complex scenes (Broda et al., 2024). Our current data did not allow us to extract comparable individual gaze divergence estimates given our stringent criteria for selected event sequences, restricted to intermediate fixation falling on inanimate objects with saccade targeting either face or object (for further details, see Figure 1). A larger dataset could provide sufficient data to derive individual estimates and use them to probe the covariance of latency and saccade velocity to test our prediction that these rest on separate mechanisms.

In both studies, sample size remains a limitation, particularly in our fMRI study, where small samples pose challenges for detecting individual differences (Marek et al., 2022). Although approaches like hyperalignment (DeYoung et al., 2022) help reduce noise and stabilize effect sizes; increased sample sizes would undoubtedly strengthen the robustness of these findings. Additionally, our second study divided the experiment into separate eye tracking and fMRI sessions, resulting in participants viewing the movie twice. While prior research suggests that neural activations (Lu et al., 2016) and gaze behavior (Dorr et al., 2010) remain largely consistent across repetitions, repetition effects (Finn et al., 2020) cannot be entirely ruled out despite counterbalancing the session order.

Another limitation is our use of an animated movie with animal characters. This choice may have influenced neural representations in the ventral cortex—an area fine-tuned to specific categories such as faces—even though previous studies indicate that face-like objects can elicit responses similar to those elicited by real faces (Hadjikhani et al., 2009). Consequently, using a wider range of stimuli, including naturalistic movies with human characters in future studies, could help determine whether the observed effects are specific to animated content or generalizable across different types of visual input.

Finally, we show that cross-brain decoding accuracy in IT is larger in fixation compared to the free-view condition. The difference between conditions served as an estimate of the neural divergence, likely induced by eye movements. Although the cross-brain decoding accuracy was high in the fixation condition, it was imperfect. One likely contributor to this variability is the presence of small eye movements, such as ocular drift and microsaccades, during the fixation condition. Although IT appears

relatively robust against the influence of microsaccades, research shows that these subtle movements can modulate activity in early visual areas (Thielen et al., 2019) and evoke responses in higher areas like V4 (Leopold, 1998). Moreover, microsaccades exhibit consistent inter-individual differences and are linked to cognitive functions such as attention and working memory (Engbert & Kliegl, 2003; Hafed & Clark, 2002). These factors may have contributed to the imperfect cross-brain decoding accuracy observed. Future studies should consider incorporating microsaccade metrics into models of gaze divergence to better clarify their impact (Ko et al., 2010; Mergenthaler & Engbert, 2010; Perquin & Bompas, 2019; Rucci et al., 2007; Turatto et al., 2007).

4.4 Future heading

4.4.1 Expanding semantic categories of dynamic content

In our second study, we examined individual differences in gaze behavior using low-level (e.g., Euclidean distances between gaze positions) and high-level (e.g., preference for faces and text) predictors. Although these measures offer valuable insights, focusing solely on faces and text limits our ability to capture the full semantic range in dynamic stimuli like movies.

Movies are rich, multifaceted stimuli in which the gaze is influenced by both bottom-up sensory features (e.g., motion; (Hutson et al., 2022)) and top-down cognitive processes such as narrative context and task goals (Çukur et al., 2013; Smith & Mital, 2013). Detailed movie annotations that incorporate narrative events and boundaries (Baldassano et al., 2017) as well as broader social context (Rubo & Gamer, 2018) may reveal how individual differences in gaze behavior are linked to higher-level cognitive functions such as narrative comprehension and memory. This approach is also supported by recent work suggesting that the occipitotemporal cortex processes objects in context rather than in isolation and with respect to behavioral goals (Bracci & De Beeck, 2023; Ritchie, 2024). While the effects of narrative elements on gaze convergence have been modest (Hutson et al., 2017), studies show that similarity in narrative interpretation is reflected in converging neural activations during viewing and recall (Chen et al., 2017; Nguyen et al., 2019; Saalasti et al., 2019). Recent evidence also suggests that individual fixation patterns relate to unique scene descriptions (Kollenda et al., 2024) and that gaze patterns are reinstated during memory recall (Nau et al., 2024). The extent to which shared gaze patterns reflect narrative comprehension and memory remains unclear. However, future research could utilize rich movie stimuli and classify them along dimensions like social context (e.g., interactive exchanges and passive co-presence), emotional valence (e.g., positive and negative interactions), and narrative role

(e.g., identifying leading versus supporting agents). This finer categorization might help to illuminate the links between individual gaze and higher cognitive functions and possibly explain why people differ so dramatically in movie appraisal (Wallisch & Whritner, 2017).

4.4.2 Towards naturalistic behavior

Our first study demonstrates that the effect of rapid saccades on faces generalizes from isolated stimulus experiments to free-viewing gaze behavior with complex stimuli. In the second study, we extended our focus to dynamic stimuli and successfully linked gaze divergence between observers to the differences in neural representations. Our two studies approached naturalistic behavior by incorporating increasingly complex stimuli in the context of free-viewing behavior, aiming for more ecologically valid results. As Holleman et al. (2020) argue, ecological validity is often ambiguously defined, making it crucial to specify the criteria used when assessing the generalizability of experimental findings. In this discussion, we consider ecological validity as the extent to which findings from controlled settings can be generalized to real-world behavior (Andrade, 2018; Lewkowicz, 2001; Schmuckler, 2001). Ecological validity can refer to an experiment's resemblance to the natural environment but also the extent to which findings generalize to real-world behavior, regardless of stimulus similarity.

To further evaluate our approach, we consider three dimensions proposed by Schmuckler (2001): (1) the nature of the stimuli, (2) the nature of the task, behavior, or response, and (3) the nature of the research context - framing them along the continuum from artificial to natural and simple to complex (Holleman et al., 2020).

For *stimuli*, while static to dynamic stimuli are increasing in complexity, e.g., movies additionally contain motion information and often narrative structure compared to static images, it is not entirely clear whether they also increase in naturality regarding their resemblance to the real world. In recent years, movie stimuli have been increasingly used to study various cognitive and perceptual processes (Feilong et al., 2021; Haxby et al., 2011; Haxby, Gobbini, et al., 2020; Haxby, Guntupalli, et al., 2020; Visconti di Oleggio Castello et al., 2020) and are often termed ecologically valid (Nastase et al., 2020; Zhang et al., 2021), largely due to their ability to evoke brain responses that are highly reproducible within and across observers (Hasson et al., 2010; Zhang et al., 2021). Although movies may remain ecologically valid for scenarios where attentional guidance is externally driven (e.g., watching a film in a cinema), their structured narrative and editing techniques impose constraints on visual exploration that differ from real-world perception, e.g., more idiosyncratic gaze behavior shown for unedited natural

video clips compared to director-cut movies (Dorr et al., 2010b; Hasson et al., 2008; Hirose, 2010). In contrast, real-world scenarios—such as sitting on a bench and observing one's surroundings—allow for unrestricted gaze behavior, leading to more variable and goal-driven eye movements. Despite some of the constraints concerning movie stimuli, they remain valuable for the study of naturalistic vision and a reasonable compromise between highly controlled experiments and unconstrained real-world behavior.

For *task behavior*, our studies included free-viewing conditions, which provide a step toward naturalistic behavior by allowing saccade and fixation dynamics to emerge freely. However, unlike real-world scenarios, where gaze behavior is shaped often by goals (e.g., searching, navigating, or interacting with others), free-viewing lacks task-driven modulations (Hayhoe & Ballard, 2005; Tatler et al., 2011). While fundamental effects like rapid orienting of gaze toward faces, as supported by our first study, likely persist beyond the lab, they may be modulated by contextual factors such as social presence (Laidlaw et al., 2011; Pasqualette & Kulke, 2023), or competing attentional demands (Hessels et al., 2023).

For *research context*, both studies were conducted in laboratory settings and could be regarded as less naturalistic by isolating an individual in a dark room or an fMRI scanner. Gibson (1979) argued that "the laboratory must be like life," and in many ways, the gap between lab and real-world settings has been narrowing. As daily life becomes increasingly structured around screens and experimental environments evolve, laboratory tasks resemble real-world behavior more than ever.

This section is not meant to disqualify the results of laboratory-based experiments. Despite the differences between the laboratory and the real world, many findings are stable and generalize well. For instance, individual gaze fixation patterns on faces in static scenes generalize to face-viewing behavior in real-world environments (Guy & Pertzov, 2023; Peterson et al., 2016). Studies exploring human gaze behavior in various tasks (e.g., making coffee) also confirm lab findings in the context of fundamental eye movement principles such as task-driven attention or predictive gaze (Hayhoe & Ballard, 2014; Fathi, Hodgins, & Rehg, 2011; Hayhoe & Ballard, 2005). However, the shift towards more naturalistic settings is highly informative and more feasible than it used to be. The increasing availability of large datasets on naturalistic gaze behavior (Fuhl et al., 2021; Greene et al., 2024; Kothari et al., 2020) presents a unique opportunity to bridge the gap between controlled experiments and real-world vision research. Additionally, advancements in mobile brain and eye-tracking (Stangl et al., 2023) now make it possible to study gaze and brain in truly natural settings.

5 Conclusion

In this thesis, two different approaches were considered: one from the perspective of a standard observer and the other focused on an individual observer, intending to draw conclusions from unconstrained gaze behavior toward complex stimuli. In the first study, we showed that faces attract gaze faster than other objects, even in complex naturalistic scenes, where multiple processes, such as concurrent processing of the current and next fixated targets, are involved. Specifically, two separate findings of higher saccade velocity and lower latency toward faces generalize from experiments with isolated stimuli to free-viewing complex scenes. Crucially, the findings that lower latency toward faces compared to saccade velocity is limited to targets that are near to the current fixation, suggest that they may depend on two distinct mechanisms. In the second study, we extended our approach from complex static images to dynamic movie stimuli. We explored the interplay of semantic information, gaze and neural patterns from an idiosyncratic perspective. We extracted individual tendencies to fixate on distinct semantic categories, along with differences in low-level oculomotor parameters, and estimated their contribution to neural representational divergence in V1, IT, and beyond. We showed individual differences in gaze toward semantic categories and the average Euclidean distance between gaze positions matter, even in the inferior temporal cortex, which was previously considered eye-movement invariant. However, these effects deserve further investigation. Distinguishing between the effects of motion information and categorical representations would help to understand how each contributes to neural divergence of observers. Additionally, deriving extended annotations of movie stimuli that consider context and observer's behavioral goals, could illuminate the effects of gaze on other cognitive capacities such as narrative understanding or memory. Finally, moving toward more naturalistic, real-world experiments could shed light on the gaze characteristics that matter for interaction with the world.

6 References

Andrade, C. (2018). Internal, External, and Ecological Validity in Research Design, Conduct, and Evaluation. *Indian Journal of Psychological Medicine*, *40*(5), 498–499. https://doi.org/10.4103/IJPSYM.IJPSYM_334_18

Andrews, T. J., & Coppola, D. M. (1999). Idiosyncratic characteristics of saccadic eye movements when viewing different visual environments. *Vision Research*, *39*(17), 2947–2953. https://doi.org/10.1016/S0042-6989(99)00019-X

Arcaro, M. J., Schade, P. F., Vincent, J. L., Ponce, C. R., & Livingstone, M. S. (2017). Seeing faces is necessary for face-domain formation. *Nature Neuroscience*, *20*(10), 1404–1412. https://doi.org/10.1038/nn.4635

Azadi, R., Lopez, E., Taubert, J., Patterson, A., & Afraz, A. (2024). Inactivation of face-selective neurons alters eye movements when free viewing faces. *Proceedings of the National Academy of Sciences*, *121*(3), 2017. https://doi.org/10.1073/pnas.2309906121

Bahill, A. T., Clark, M. R., & Stark, L. (1975). The main sequence, a tool for studying human eye movements. *Mathematical Biosciences*, *24*(3–4), 191–204. https://doi.org/10.1016/0025-5564(75)90075-9

Baldassano, C., Chen, J., Zadbood, A., Pillow, J. W., Hasson, U., & Norman, K. A. (2017). Discovering Event Structure in Continuous Narrative Perception and Memory. *Neuron*, *95*(3), 709-721.e5. https://doi.org/10.1016/j.neuron.2017.06.041

Bargary, G., Bosten, J. M., Goodbourn, P. T., Lawrance-Owen, A. J., Hogg, R. E., & Mollon, J. D. (2017). Individual differences in human eye movements: An oculomotor signature? *Vision Research*, *141*, 157–169. https://doi.org/10.1016/j.visres.2017.03.001

Beltramo, R., & Scanziani, M. (2019). A collicular visual cortex: Neocortical space for an ancient midbrain visual structure. *Science*, *363*(6422), 64–69. https://doi.org/10.1126/science.aau7052

Benson, N. C., Butt, O. H., Datta, R., Radoeva, P. D., Brainard, D. H., & Aguirre, G. K. (2012). The retinotopic organization of striate cortex is well predicted by surface topology. *Current Biology*, *22*(21), 2081–2085. https://doi.org/10.1016/j.cub.2012.09.014

Benson, N. C., Yoon, J. M. D., Forenzo, D., Engel, S. A., Kay, K. N., & Winawer, J. (2022). Variability of the Surface Area of the V1, V2, and V3 Maps in a Large Sample of Human Observers. *Journal of Neuroscience*, *42*(46), 8629–8646. https://doi.org/10.1523/JNEUROSCI.0690-21.2022

Borovska, P., & de Haas, B. (2023). Faces in scenes attract rapid saccades. *Journal of Vision*, *23*(8), 1–15. https://doi.org/10.1167/jov.23.8.11

Borovska, P., & de Haas, B. (2024). Individual gaze shapes diverging neural representations. *Proceedings of the National Academy of Sciences*, *121*(36), 2017. https://doi.org/10.1073/pnas.2405602121

Boucart, M., & Thorpe, S. J. (2016). *Finding faces , animals , and vehicles in far peripheral vision*. *16*, 1–13. https://doi.org/10.1167/16.2.10.doi

Bracci, S., & De Beeck, H. P. O. (2023). Understanding Human Object Vision: A Picture Is Worth a Thousand Representations. *Annual Review of Psychology*, *74*, 113–135. https://doi.org/10.1146/annurev-psych-032720-041031

Bracci, S., Ritchie, J. B., & de Beeck, H. O. (2017). On the partnership between neural representations of object categories and visual features in the ventral visual pathway. *Neuropsychologia*, *105*(October 2016), 153–164. https://doi.org/10.1016/j.neuropsychologia.2017.06.010

Broda, M. D., Borovska, P., & Haas, B. De. (2024). Individual differences in face salience and rapid face saccades. *Journal of Vision*, *24*(6), 1–24. https://doi.org/https://doi.org/10.1167/jov.24.6.16

Broda, M. D., & de Haas, B. (2022). Individual fixation tendencies in person viewing generalize from images to videos. *I-Perception*, *13*(6). https://doi.org/10.1177/20416695221128844

Broda, M. D., Haddad, T., & de Haas, B. (2023). Quick, eyes! Isolated upper face regions but not artificial features elicit rapid saccades. *Journal of Vision*, *23*(2), 1–9. https://doi.org/10.1167/jov.23.2.5

Broers, N., Bainbridge, W. A., Michel, R., Balestrieri, E., & Busch, N. A. (2022). The extent and specificity of visual exploration determines the formation of recollected memories in complex scenes. *Journal of Vision*, *22*(11), 1–12. https://doi.org/10.1167/jov.22.11.9

Busch, E. L., Rapuano, K. M., Anderson, K. M., Rosenberg, M. D., Watts, R., Casey, B. J., Haxby, J. V., & Feilong, M. (2024). Dissociation of Reliability, Heritability, and Predictivity in Coarse- and Fine-Scale Functional Connectomes during Development. *Journal of Neuroscience*, *44*(6), 1–16. https://doi.org/10.1523/JNEUROSCI.0735-23.2023

Cerf, M., Paxon Frady, E., & Koch, C. (2009). Faces and text attract gaze independent of the task: Experimental data and computer model. *Journal of Vision*, *9*(12), 1–15. https://doi.org/10.1167/9.12.1

Charest, I., & Kriegeskorte, N. (2015). The brain of the beholder: honouring individual representational idiosyncrasies. *Language, Cognition and Neuroscience*, *30*(4), 367–379. https://doi.org/10.1080/23273798.2014.1002505

Chen, J., Leong, Y. C., Honey, C. J., Yong, C. H., Norman, K. A., & Hasson, U. (2017). Shared memories reveal shared structure in neural activity across individuals. *Nature Neuroscience*, *20*(1), 115–125. https://doi.org/10.1038/nn.4450

Coutrot, A., & Guyader, N. (2014). How saliency, faces, and sound influence gaze in dynamic social scenes. *Journal of Vision*, *14*(8), 1–17. https://doi.org/10.1167/14.8.5

Crouzet, S. M. (2010). Fast saccades toward faces: Face detection in just 100 ms. *Journal of Vision*, *10*(4), 1–17. https://doi.org/10.1167/10.4.16

Çukur, T., Nishimoto, S., Huth, A. G., & Gallant, J. L. (2013). Attention during natural vision warps semantic representation across the human brain. *Nature Neuroscience*, *16*(6), 763–770. https://doi.org/10.1038/nn.3381

Cutts, S. A., Faskowitz, J., Betzel, R. F., & Sporns, O. (2023). Uncovering individual differences in fine-scale dynamics of functional connectivity. *Cerebral Cortex*, *33*(5), 2375–2394. https://doi.org/10.1093/cercor/bhac214

Damiano, C., & Walther, D. B. (2019). Distinct roles of eye movements during memory encoding and retrieval. *Cognition*, *184*(December 2018), 119–129. https://doi.org/10.1016/j.cognition.2018.12.014

De Haas, B., Iakovidis, A. L., Schwarzkopf, D. S., & Gegenfurtner, K. R. (2019). Individual differences in visual salience vary along semantic dimensions. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(24), 11687–11692. https://doi.org/10.1073/pnas.1820553116

DeYoung, C. G., Sassenberg, T. A., Abend, R., Allen, T. A., Beaty, R. E., Bellgrove, M. A., Blain, S. D., Bzdok, D., Chavez, R. S., Engel, S. A., FFeilong, M., Fornito, A., Genc, E., Goghari, V., Grazioplene, R. G., Hanson, J. L., Haxb, J. V, Hilger, K., Homan, P., … Wacker, J. (2022). Preprint: Reproducible between-pwerson brain-behavior associations do not alwasy require thousands of individuals. *PsyArXiv*, *5*(1), 47–55.

Dorr, M., Gegenfurtner, K. R., & Barth, E. (2009). The contribution of low-level features at the centre of gaze to saccade target selection. *Vision Research*, *49*(24), 2918–2926. https://doi.org/10.1016/j.visres.2009.09.007

Dorr, M., Martinetz, T., Gegenfurtner, K. R., & Barth, E. (2010a). Variability of eye movements when viewing dynamic natural scenes. *Journal of Vision*, *10*(10), 1–17. https://doi.org/10.1167/10.10.28

Dorr, M., Martinetz, T., Gegenfurtner, K. R., & Barth, E. (2010b). Variability of eye movements when viewing dynamic natural scenes. *Journal of Vision*, *10*(10), 1–17. https://doi.org/10.1167/10.10.28

Dougherty, R. F., Koch, V. M., Brewer, A. A., Fischer, B., Modersitzki, J., & Wandell, B. A. (2003). Visual field representations and locations of visual areas v1/2/3 in human visual cortex. *Journal of Vision*, *3*(10), 586–598. https://doi.org/10.1167/3.10.1

Einhäuser, W., Atzert, C., & Nuthmann, A. (2020). Fixation durations in natural scene viewing are guided by peripheral scene content. *Journal of Vision*, *20*(4), 1–15. https://doi.org/10.1167/jov.20.4.15

Engbert, R., & Kliegl, R. (2003). Microsaccades uncover the orientation of covert attention. *Vision Research*, *43*(9), 1035–1045. https://doi.org/10.1016/S0042-6989(03)00084-1

Engmann, S., Hart, B. M. 't, Sieren, T., Onat, S., König, P., & Einhäuser, W. (2009). Saliency on a natural scene background: Effects of color and luminance contrast add linearly. *Attention, Perception, & Psychophysics*, *71*(6), 1337–1352. https://doi.org/10.3758/APP.71.6.1337

Fehlmann, B., Coynel, D., Schicktanz, N., Milnik, A., Gschwind, L., Hofmann, P., Papassotiropoulos, A., & de Quervain, D. J.-F. (2020). Visual Exploration at Higher Fixation Frequency Increases Subsequent Memory Recall. *Cerebral Cortex Communications*, *1*(1), 1–14. https://doi.org/10.1093/texcom/tgaa032

Feilong, M., Nastase, S. A., Guntupalli, J. S., & Haxby, J. V. (2018). Reliable individual differences in fine-grained cortical functional architecture. *NeuroImage*, *183*(September), 375–386. https://doi.org/10.1016/j.neuroimage.2018.08.029

Feilong, M., Nastase, S. A., Jiahui, G., Halchenko, Y. O., Gobbini, M. I., & Haxby, J. V. (2022). *The Individualized Neural Tuning Model: Precise and generalizable cartography of functional architecture in individual brains*. *1*(March). https://doi.org/10.32470/ccn.2022.1216-0

Feilong, M., Swaroop Guntupalli, J., & Haxby, J. V. (2021). The neural basis of intelligence in fine-grained cortical topographies. *ELife*, *10*, 1–33. https://doi.org/10.7554/eLife.64058

Finn, E. S., & Bandettini, P. A. (2021). Movie-watching outperforms rest for functional connectivity-based prediction of behavior. *NeuroImage*, *235*(March), 117963. https://doi.org/10.1016/j.neuroimage.2021.117963

Finn, E. S., Glerean, E., Khojandi, A. Y., Nielson, D., Molfese, P. J., Handwerker, D. A., & Bandettini, P. A. (2020). Idiosynchrony: From shared responses to individual differences during naturalistic neuroimaging. *NeuroImage*, *215*(April), 116828. https://doi.org/10.1016/j.neuroimage.2020.116828

Finzi, D., Gomez, J., Nordt, M., Rezai, A. A., Poltoratski, S., & Grill-Spector, K. (2021). Differential spatial computations in ventral and lateral face-selective regions are scaffolded by structural connections. *Nature Communications*, *12*(1), 2278. https://doi.org/10.1038/s41467-021-22524-2

Foulsham, T., Cheng, J. T., Tracy, J. L., Henrich, J., & Kingstone, A. (2010). Gaze allocation in a dynamic situation: Effects of social status and speaking. *Cognition*, *117*(3), 319–331. https://doi.org/https://doi.org/10.1016/J.COGNITION.2010.09.003

Fuhl, W., Kasneci, G., & Kasneci, E. (2021). TEyeD: Over 20 Million Real-World Eye Images with Pupil, Eyelid, and Iris 2D and 3D Segmentations, 2D and 3D Landmarks, 3D Eyeball, Gaze

Vector, and Eye Movement Types. *2021 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, 367–375. https://doi.org/10.1109/ISMAR52148.2021.00053

Furman, O., Dorfman, N., Hasson, U., Davachi, L., & Dudai, Y. (2007). They saw a movie: long-term memory for an extended audiovisual narrative. *Learning & Memory (Cold Spring Harbor, N.Y.)*, *14*(6), 457–467. https://doi.org/10.1101/lm.550407

Geerligs, L., Rubinov, M., Tyler, L. K., Brayne, C., Bullmore, E. T., Calder, A. C., Cusack, R., Dalgleish, T., Duncan, J., Henson, R. N., Matthews, F. E., Marslen-Wilson, W. D., Rowe, J. B., Shafto, M. A., Campbell, K., Cheung, T., Davis, S., Geerligs, L., Kievit, R., … Henson, R. N. (2015). State and trait components of functional connectivity: Individual differences vary with mental state. *Journal of Neuroscience*, *35*(41), 13949–13961. https://doi.org/10.1523/JNEUROSCI.1324-15.2015

Gibson, J. J. (1979). *The ecological approach to visual perception*. Houghton, Mifflin and Company. https://doi.org/10.4324/9781315740218

Greene, M. R., Balas, B. J., Lescroart, M. D., MacNeilage, P. R., Hart, J. A., Binaee, K., Hausamann, P. A., Mezile, R., Shankar, B., Sinnott, C. B., Capurro, K., Halow, S., Howe, H., Josyula, M., Li, A., Mieses, A., Mohamed, A., Nudnou, I., Parkhill, E., … Weissmann, E. (2024). The visual experience dataset: Over 200 recorded hours of integrated eye movement, odometry, and egocentric video. *Journal of Vision*, *24*(11), 6. https://doi.org/10.1167/jov.24.11.6

Grill-Spector, K., & Weiner, K. S. (2014). The functional architecture of the ventral temporal cortex and its role in categorization. *Nat Rev Neurosci*, *15*(8), 536–548. https://doi.org/10.1038/nrn3747

Grill-Spector, K., Weiner, K. S., Kay, K., & Gomez, J. (2017). The Functional Neuroanatomy of Human Face Perception. *Annual Review of Vision Science*, *3*, 167–196. https://doi.org/10.1146/annurev-vision-102016-061214

Groen, I. I. A., Silson, E. H., & Baker, C. I. (2017). Contributions of low- and high-level properties to neural processing of visual scenes in the human brain. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *372*(1714), 20160102. https://doi.org/10.1098/rstb.2016.0102

Grosbras, M. H., Laird, A. R., & Paus, T. (2005). Cortical regions involved in eye movements, shifts of attention, and gaze perception. *Human Brain Mapping*, *25*(1), 140–154. https://doi.org/10.1002/hbm.20145

Güçlü, U., & van Gerven, M. A. J. (2017). Increasingly complex representations of natural movies across the dorsal stream are shared between subjects. *NeuroImage*, *145*, 329–336. https://doi.org/10.1016/j.neuroimage.2015.12.036

Guntupalli, J. S., Hanke, M., Halchenko, Y. O., Connolly, A. C., Ramadge, P. J., & Haxby, J. V. (2016). A Model of Representational Spaces in Human Cortex. *Cerebral Cortex*, *26*(6), 2919–2934. https://doi.org/10.1093/cercor/bhw068

Guntupalli, J. S., Wheeler, K. G., & Gobbini, M. I. (2017). Disentangling the Representation of Identity from Head View Along the Human Face Processing Pathway. *Cerebral Cortex*, *27*(1), 46–53. https://doi.org/10.1093/cercor/bhw344

Guo, K., Mahmoodi, S., Robertson, R. G., & Young, M. P. (2006). Longer fixation duration while viewing face images. *Experimental Brain Research*, *171*(1), 91–98. https://doi.org/10.1007/s00221-005-0248-y

Guy, N., & Pertzov, Y. (2023). The robustness of individual differences in gaze preferences toward faces and eyes across face-to-face experimental designs and its relation to social anxiety. *Journal of Vision*, *23*(5), 1–13. https://doi.org/10.1167/jov.23.5.15

Hadjikhani, N., Kveraga, K., Naik, P., & Ahlfors, S. P. (2009). Early (M170) activation of face-specific cortex by face-like objects. *NeuroReport*, *20*(4), 403–407. https://doi.org/10.1097/WNR.0b013e328325a8e1

Hafed, Z. M., & Clark, J. J. (2002). Microsaccades as an overt measure of covert attention shifts. *Vision Research*, *42*(22), 2533–2545. https://doi.org/10.1016/S0042-6989(02)00263-8

Harel, J., Koch, C., & Perona, P. (2007). Graph-based visual saliency. *Advances in Neural Information Processing Systems*, *19*, 545–552. https://doi.org/10.7551/mitpress/7503.003.0073

Hasson, U., Landesman, O., Knappmeyer, B., Vallines, I., Rubin, N., & Heeger, D. J. (2008). *Neurocinematics: The Neuroscience of Film*. *2*(1), 1–26. http://ohadlandesman.com/pdf/Neurocinematics-Projections2008.pdf

Hasson, U., Levy, I., Behrmann, M., Hendler, T., & Malach, R. (2002). Eccentricity bias as an organizing principle for human high-order object areas. *Neuron*, *34*(3), 479–490. https://doi.org/10.1016/S0896-6273(02)00662-1

Hasson, U., Malach, R., & Heeger, D. J. (2010). Reliability of cortical activity during natural stimulation. *Trends in Cognitive Sciences*, *14*, 40–48. https://api.semanticscholar.org/CorpusID:6229276

Hasson, U., Nir, Y., Levy, I., Galit Fuhrmann, & Malach, R. (2004). Intersubject Synchronization of Cortical Activity during Natural Vision. *Science*, *303*(5664), 1634–40.

Haxby, J. V., Gobbini, M. I., & Nastase, S. A. (2020). Naturalistic stimuli reveal a dominant role for agentic action in visual representation. *NeuroImage*, *216*(August 2019), 116561. https://doi.org/10.1016/j.neuroimage.2020.116561

Haxby, J. V., Guntupalli, J. S., Connolly, A. C., Halchenko, Y. O., Conroy, B. R., Gobbini, M. I., Hanke, M., & Ramadge, P. J. (2011). A common, high-dimensional model of the representational space in human ventral temporal cortex. *Neuron*, *72*(2), 404–416. https://doi.org/10.1016/j.neuron.2011.08.026

Haxby, J. V., Guntupalli, J. S., Nastase, S. A., & Feilong, M. (2020). Hyperalignment: Modeling shared information encoded in idiosyncratic cortical topographies. *ELife*, *9*, 1–26. https://doi.org/10.7554/eLife.56601

Hayhoe, M., & Ballard, D. (2005). Eye movements in natural behavior. *Trends in Cognitive Sciences*, *9*(4), 188–194. https://doi.org/10.1016/j.tics.2005.02.009

Hayhoe, M. M., McKinney, T., Chajka, K., & Pelz, J. B. (2012). Predictive eye movements in natural vision. *Experimental Brain Research*, *217*(1), 125–136. https://doi.org/10.1007/s00221-011-2979-2

Herwig, A., & Schneider, W. X. (2014). Predicting object features across saccades: Evidence from object recognition and visual search. *Journal of Experimental Psychology: General*, *143*(5), 1903–1922. https://doi.org/10.1037/a0036781

Hessels, R. S., Teunisse, M. K., Niehorster, D. C., Nyström, M., Benjamins, J. S., Senju, A., & Hooge, I. T. C. (2023). Task-related gaze behaviour in face-to-face dyadic collaboration: Toward an interactive theory? *Visual Cognition*, *31*(4), 291–313. https://doi.org/10.1080/13506285.2023.2250507

Himmelberg, M. M., Winawer, J., & Carrasco, M. (2022). Linking individual differences in human primary visual cortex to contrast sensitivity around the visual field. *Nature Communications*, *13*(1), 3309. https://doi.org/10.1038/s41467-022-31041-9

Hirose, Y. (2010). Perception and memory across viewpoint changes in moving images. *Journal of Vision*, *10*(4), 1–19. https://doi.org/10.1167/10.4.2

Holleman, G. A., Hooge, I. T. C., Kemner, C., & Hessels, R. S. (2020). The 'Real-World Approach' and Its Problems: A Critique of the Term Ecological Validity. *Frontiers in Psychology*, *11*(April), 1–12. https://doi.org/10.3389/fpsyg.2020.00721

Huber-Huber, C., Buonocore, A., & Melcher, D. (2021). The extrafoveal preview paradigm as a measure of predictive, active sampling in visual perception. *Journal of Vision*, *21*(7), 1–23. https://doi.org/10.1167/JOV.21.7.12

Huth, A. G., Nishimoto, S., Vu, A. T., & Gallant, J. L. (2012). A Continuous Semantic Space Describes the Representation of Thousands of Object and Action Categories across the Human Brain. *Neuron*, *76*(6), 1210–1224. https://doi.org/10.1016/j.neuron.2012.10.014

Hutson, J. P., Chandran, P., Magliano, J. P., Smith, T. J., & Loschky, L. C. (2022). Narrative Comprehension Guides Eye Movements in the Absence of Motion. *Cognitive Science*, *46*(5). https://doi.org/10.1111/cogs.13131

Hutson, J. P., Smith, T. J., Magliano, J. P., & Loschky, L. C. (2017). What is the role of the film viewer? The effects of narrative comprehension and viewing task on gaze control in film.

Cognitive Research: Principles and Implications, *2*(1). https://doi.org/10.1186/s41235-017-0080-5

Itti, L., & Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research*, *40*(10–12), 1489–1506. https://doi.org/10.1016/S0042-6989(99)00163-7

Jääskeläinen, I. P., Sams, M., Glerean, E., & Ahveninen, J. (2021). Movies and narratives as naturalistic stimuli in neuroimaging. *NeuroImage*, *224*(June 2020). https://doi.org/10.1016/j.neuroimage.2020.117445

Jayaraman, S., & Smith, L. B. (2019). Faces in early visual environments are persistent not just frequent. *Vision Research*, *157*(October 2017), 213–221. https://doi.org/10.1016/j.visres.2018.05.005

Jiahui, G., Feilong, M., Nastase, S. A., Haxby, J. V., & Gobbini, M. I. (2023). Cross-movie prediction of individualized functional topography. *ELife*, *12*, 1–17. https://doi.org/10.7554/eLife.86037

Jiahui, G., Feilong, M., Visconti di Oleggio Castello, M., Guntupalli, J. S., Chauhan, V., Haxby, J. V., & Gobbini, M. I. (2020). Predicting individual face-selective topography using naturalistic stimuli. *NeuroImage*, *216*(December), 116458. https://doi.org/10.1016/j.neuroimage.2019.116458

Kanwisher, N. (2010). Functional specificity in the human brain: A window into the functional architecture of the mind. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(25), 11163–11170. https://doi.org/10.1073/pnas.1005062107

Kauffmann, L., Peyrin, C., Chauvin, A., Entzmann, L., Breuil, C., & Guyader, N. (2019). Face perception influences the programming of eye movements. *Scientific Reports*, *9*(1), 1–14. https://doi.org/10.1038/s41598-018-36510-0

Kay, K. N., Naselaris, T., Prenger, R. J., & Gallant, J. L. (2008). Identifying natural images from human brain activity. *Nature*, *452*(7185), 352–355. https://doi.org/10.1038/nature06713

Ko, H., Poletti, M., & Rucci, M. (2010). Microsaccades precisely relocate gaze in a high visual acuity task. *Nature Neuroscience*, *13*(12), 1549–1553. https://doi.org/10.1038/nn.2663

Kollenda, D., Reher, A. V., & de Haas, B. (2024). Individual gaze predicts individual scene descriptions. *PsyArxiv*. https://doi.org/https://doi.org/10.31234/osf.io/nx7jy

Kothari, R., Yang, Z., Kanan, C., Bailey, R., Pelz, J. B., & Diaz, G. J. (2020). Gaze-in-wild: A dataset for studying eye and head coordination in everyday activities. *Scientific Reports*, *10*(1), 1–18. https://doi.org/10.1038/s41598-020-59251-5

Kravitz, D. J., Saleem, K. S., Baker, C. I., Ungerleider, L. G., & Mishkin, M. (2013). The ventral visual pathway: An expanded neural framework for the processing of object quality. *Trends in Cognitive Sciences*, *17*(1), 26–49. https://doi.org/10.1016/j.tics.2012.10.011

Kriegeskorte, N., & Kievit, R. A. (2013). Representational geometry: Integrating cognition, computation, and the brain. *Trends in Cognitive Sciences*, *17*(8), 401–412. https://doi.org/10.1016/j.tics.2013.06.007

Kriegeskorte, N., Mur, M., & Bandettini, P. (2008). Representational similarity analysis - connecting the branches of systems neuroscience. *Frontiers in Systems Neuroscience*, *2*(NOV), 1–28. https://doi.org/10.3389/neuro.06.004.2008

Krishna, B. S., Ipata, A. E., Bisley, J. W., Gottlieb, J., & Goldberg, M. E. (2014). Extrafoveal preview benefit during free-viewing visual search in the monkey. *Journal of Vision*, *14*(1), 6–6. https://doi.org/10.1167/14.1.6

Kucharský, Š., van Renswoude, D., Raijmakers, M., & Visser, I. (2021). WALD-EM: Wald accumulation for locations and durations of eye movements. *Psychological Review*, *128*(4), 667–689. https://doi.org/10.1037/rev0000292

Kümmerer, M., & Bethge, M. (2021). *State-of-the-Art in Human Scanpath Prediction*. http://arxiv.org/abs/2102.12239

Laidlaw, K. E. W., Foulsham, T., Kuhn, G., & Kingstone, A. (2011). Potential social interactions are important to social attention. *Proceedings of the National Academy of Sciences*, *108*(14), 5548–5553. https://doi.org/10.1073/pnas.1017022108

Laumann, T. O., Gordon, E. M., Adeyemo, B., Snyder, A. Z., Joo, S. J., Chen, M. Y., Gilmore, A. W., McDermott, K. B., Nelson, S. M., Dosenbach, N. U. F., Schlaggar, B. L., Mumford, J. A., Poldrack, R. A., & Petersen, S. E. (2015). Functional System and Areal Organization of a Highly Sampled Individual Human Brain. *Neuron*, *87*(3), 657–670. https://doi.org/10.1016/j.neuron.2015.06.037

Levy, I., Hasson, U., Avidan, G., Hendler, T., & Malach, R. (2001). Center-periphery organization of human object areas. *Nature Neuroscience*, *4*(5), 533–539. https://doi.org/10.1038/87490

Lewkowicz, D. J. (2001). The Concept of Ecological Validity: What are Its Limitations and is It Bad to Be Invalid? *Infancy*, *2*(4), 437–450. https://doi.org/10.1207/S15327078IN0204_03

Liu, Z.-X., Rosenbaum, R. S., & Ryan, J. D. (2020). Restricting Visual Exploration Directly Impedes Neural Activity, Functional Connectivity, and Memory. *Cerebral Cortex Communications*, *1*(1), 1–15. https://doi.org/10.1093/texcom/tgaa054

Ludwig, C. J. H., Davies, J. R., & Eckstein, M. P. (2014). Foveal analysis and peripheral selection during active visual sampling. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(2), 291–299. https://doi.org/10.1073/pnas.1313553111

Lu, K. H., Hung, S. C., Wen, H., Marussich, L., & Liu, Z. (2016). Influences of high-level features, gaze, and scene transitions on the reliability of BOLD responses to natural movie stimuli. *PLoS ONE*, *11*(8), 1–19. https://doi.org/10.1371/journal.pone.0161797

Marek, S., Tervo-Clemmens, B., Calabro, F. J., Montez, D. F., Kay, B. P., Hatoum, A. S., Donohue, M. R., Foran, W., Miller, R. L., Hendrickson, T. J., Malone, S. M., Kandala, S., Feczko, E., Miranda-Dominguez, O., Graham, A. M., Earl, E. A., Perrone, A. J., Cordova, M., Doyle, O., … Dosenbach, N. U. F. (2022). Reproducible brain-wide association studies require thousands of individuals. *Nature*, *603*(7902), 654–660. https://doi.org/10.1038/s41586-022-04492-9

Margalit, E., Lee, H., Finzi, D., DiCarlo, J. J., Grill-Spector, K., & Yamins, D. L. K. (2023). A Unifying Principle for the Functional Organization of Visual Cortex. *BioRxiv*, 2023.05.18.541361. https://www.biorxiv.org/content/10.1101/2023.05.18.541361v1%0Ahttps://www.biorxiv.org/content/10.1101/2023.05.18.541361v1.abstract

Martin, J. G., Davis, C. E., Riesenhuber, M., & Thorpe, S. J. (2018). Zapping 500 faces in less than 100 seconds : Evidence for extremely fast and sustained continuous visual search. *Scientific Reports*, *November 2017*, 1–12. https://doi.org/10.1038/s41598-018-30245-8

Masarwa, S., Kreichman, O., & Gilaie-Dotan, S. (2022). Larger images are better remembered during naturalistic encoding. *Proceedings of the National Academy of Sciences of the United States of America*, *119*(4). https://doi.org/10.1073/pnas.2119614119

Mergenthaler, K., & Engbert, R. (2010). Microsaccades are different from saccades in scene perception. *Experimental Brain Research*, *203*(4), 753–757. https://doi.org/10.1007/s00221-010-2272-9

Meshulam, M., Hasenfratz, L., Hillman, H., Liu, Y.-F., Nguyen, M., Norman, K. A., & Hasson, U. (2021). Neural alignment predicts learning outcomes in students taking an introduction to computer science course. *Nature Communications*, *12*(1), 1922. https://doi.org/10.1038/s41467-021-22202-3

Mollon, J. D., Bosten, J. M., Peterzell, D. H., & Webster, M. A. (2017). Individual differences in visual science: What can be learned and what is good experimental practice? *Vision Research*, *141*(November), 4–15. https://doi.org/10.1016/j.visres.2017.11.001

Nastase, S. A., Goldstein, A., & Hasson, U. (2020). Keep it real: rethinking the primacy of experimental control in cognitive neuroscience. *NeuroImage*, *222*(December 2019), 117254. https://doi.org/10.1016/j.neuroimage.2020.117254

Nau, M., Greene, A., Tarder-stoll, H., Lossio-ventura, J. A., Pereira, F., Chen, J., Baldassano, C., & Baker, C. I. (2024). *Neural and behavioral reinstatement jointly reflect retrieval of narrative events*.

Nguyen, M. N., Matsumoto, J., Hori, E., Maior, R. S., Tomaz, C., Tran, A. H., Ono, T., & Nishijo, H. (2014). Neuronal responses to face-like and facial stimuli in the monkey superior colliculus. *Frontiers in Behavioral Neuroscience*, *8*. https://doi.org/10.3389/fnbeh.2014.00085

Nguyen, M., Vanderwal, T., & Hasson, U. (2019). Shared understanding of narratives is correlated with shared neural responses. *NeuroImage*, *184*(August 2018), 161–170. https://doi.org/10.1016/j.neuroimage.2018.09.010

Nishimoto, S., Huth, A. G., Bilenko, N. Y., & Gallant, J. L. (2017). Eye movement-invariant representations in the human visual system. *Journal of Vision*, *17*(1), 1–10. https://doi.org/10.1167/17.1.11

Nuthmann, A. (2017). Fixation durations in scene viewing: Modeling the effects of local image features, oculomotor parameters, and task. *Psychonomic Bulletin and Review*, *24*(2), 370–392. https://doi.org/10.3758/s13423-016-1124-4

Osterbrink, C., & Herwig, A. (2021). Prediction of complex stimuli across saccades. *Journal of Vision*, *21*(2), 1–15. https://doi.org/10.1167/jov.21.2.10

Parker, P. R. L., Martins, D. M., Leonard, E. S. P., Casey, N. M., Sharp, S. L., Abe, E. T. T., Smear, M. C., Yates, J. L., Mitchell, J. F., & Niell, C. M. (2023). A dynamic sequence of visual processing initiated by gaze shifts. *Nature Neuroscience*, *26*(12), 2192–2202. https://doi.org/10.1038/s41593-023-01481-7

Park, J., Soucy, E., Segawa, J., Mair, R., & Konkle, T. (2023). Ultra-wide angle neuroimaging : insights into immersive scene representation. *BioRxiv*, 1–23.

Parvizi, J., Jacques, C., Foster, B. L., Withoft, N., Rangarajan, V., Weiner, K. S., & Grill-Spector, K. (2012). Electrical stimulation of Human Fusiform face-selective regions distorts face perception. *Journal of Neuroscience*, *32*(43), 14915–14920. https://doi.org/10.1523/JNEUROSCI.2609-12.2012

Pasqualette, L., & Kulke, L. (2023). Effects of emotional content on social inhibition of gaze in live social and non-social situations. *Scientific Reports*, *13*(1), 14151. https://doi.org/10.1038/s41598-023-41154-w

Perquin, M. N., & Bompas, A. (2019). Reliability and correlates of intra-individual variability in the oculomotor system. *Journal of Eye Movement Research*, *12*(6). https://doi.org/10.16910/jemr.12.6.11

Peterson, M. F., & Eckstein, M. P. (2013). Individual Differences in Eye Movements During Face Identification Reflect Observer-Specific Optimal Points of Fixation. *Psychological Science*, *24*(7), 1216–1225. https://doi.org/10.1177/0956797612471684

Peterson, M. F., Lin, J., Zaun, I., & Kanwisher, N. (2016). Individual differences in face-looking behavior generalize from the lab to the world. *Journal of Vision*, *16*(7). https://doi.org/10.1167/16.7.12

Piasini, E., Soltuzu, L., Muratore, P., Caramellino, R., Vinken, K., Op de Beeck, H., Balasubramanian, V., & Zoccolan, D. (2021). Temporal stability of stimulus representation increases along rodent

visual cortical hierarchies. *Nature Communications*, *12*(1), 4448. https://doi.org/10.1038/s41467-021-24456-3

Qiu, A., Rosenau, B. J., Greenberg, A. S., Hurdal, M. K., Barta, P., Yantis, S., & Miller, M. I. (2006). Estimating linear cortical magnification in human primary visual cortex via dynamic programming. *NeuroImage*, *31*(1), 125–138. https://doi.org/10.1016/j.neuroimage.2005.11.049

Reppert, T. R., Lempert, K. M., Glimcher, P. W., & Shadmehr, R. (2015). Modulation of saccade vigor during value–based decision making. *Journal of Neuroscience*, *35*(46), 15369–15378. https://doi.org/10.1523/JNEUROSCI.2621-15.2015

Ritchie, J. B. (2024). *Rethinking category-selectivity in human visual cortex*. 1–23. https://doi.org/https://doi.org/10.48550/arXiv.2411.08251

Rosenholtz, R. (2016). Capabilities and Limitations of Peripheral Vision. *Annual Review of Vision Science*, *2*, 437–457. https://doi.org/10.1146/annurev-vision-082114-035733

Roth, N., Rolfs, M., Hellwich, O., & Obermayer, K. (2023). Objects guide human gaze behavior in dynamic real-world scenes. In *PLoS Computational Biology* (Vol. 19, Issue 10 October). https://doi.org/10.1371/journal.pcbi.1011512

Rubo, M., & Gamer, M. (2018). Social content and emotional valence modulate gaze fixations in dynamic scenes. *Scientific Reports*, *8*(1), 1–12. https://doi.org/10.1038/s41598-018-22127-w

Rucci, M., Iovin, R., Poletti, M., & Santini, F. (2007). Miniature eye movements enhance fine spatial detail. *Nature*, *447*(7146), 852–855. https://doi.org/10.1038/nature05866

Russ, B. E., & Leopold, D. A. (2015). Functional MRI mapping of dynamic visual features during natural viewing in the macaque. *NeuroImage*, *109*, 84–94. https://doi.org/10.1016/j.neuroimage.2015.01.012

Rust, N. C., & Movshon, J. A. (2005). In praise of artifice. *Nature Neuroscience*, *8*(12), 1647–1650. https://doi.org/10.1038/nn1606

Ryan, J. D., & Shen, K. (2020). The eyes are a window into memory. *Current Opinion in Behavioral Sciences*, *32*, 1–6. https://doi.org/10.1016/j.cobeha.2019.12.014

Saalasti, S., Alho, J., Bar, M., Glerean, E., Honkela, T., Kauppila, M., Sams, M., & Jääskeläinen, I. P. (2019). Inferior parietal lobule and early visual areas support elicitation of individualized meanings during narrative listening. *Brain and Behavior*, *9*(5). https://doi.org/10.1002/brb3.1288

Schmuckler, M. A. (2001). What Is Ecological Validity? A Dimensional Analysis. *Infancy*, *2*(4), 419–436. https://doi.org/10.1207/S15327078IN0204_02

Schwarzlose, R. F., Swisher, J. D., Dang, S., & Kanwisher, N. (2008). The distribution of category and location information across object-selective regions in human visual cortex. *Proceedings of the National Academy of Sciences*, *105*(11), 4447–4452. https://doi.org/10.1073/pnas.0800431105

Schwetlick, L., Rothkegel, L. O. M., Trukenbrod, H. A., & Engbert, R. (2020). Modeling the effects of perisaccadic attention on gaze statistics during scene viewing. *Communications Biology*, *3*(1), 1–11. https://doi.org/10.1038/s42003-020-01429-8

Sha, L., & Haxby, J. V. (2015). *The Animacy Continuum in the Human Ventral Vision Pathway Long*. 139. https://doi.org/10.1162/jocn_a_00733

Silson, E. H., Chan, A. W.-Y., Reynolds, R. C., Kravitz, D. J., & Baker, C. I. (2015). A Retinotopic Basis for the Division of High-Level Scene Processing between Lateral and Ventral Human Occipitotemporal Cortex. *Journal of Neuroscience*, *35*(34), 11921–11935. https://doi.org/10.1523/JNEUROSCI.0137-15.2015

Silson, E. H., Groen, I. I. A., & Baker, C. I. (2022). Direct comparison of contralateral bias and face/scene selectivity in human occipitotemporal cortex. *Brain Structure and Function*, *227*(4), 1405–1421. https://doi.org/10.1007/s00429-021-02411-8

Smith, T. J., & Mital, P. K. (2013). Attentional synchrony and the influence of viewing task on gaze behavior in static and dynamic scenes. *Journal of Vision*, *13*(8), 1–24. https://doi.org/10.1167/13.8.16

Soares, S. C., Maior, R. S., Isbell, L. A., Tomaz, C., & Nishijo, H. (2017). Fast Detector/First Responder: Interactions between the Superior Colliculus-Pulvinar Pathway and Stimuli Relevant to Primates. *Frontiers in Neuroscience*, *11*. https://doi.org/10.3389/fnins.2017.00067

Stangl, M., Maoz, S. L., & Suthana, N. (2023). Mobile cognition: imaging the human brain in the 'real world.' *Nature Reviews Neuroscience*, *24*(6), 347–362. https://doi.org/10.1038/s41583-023-00692-y

Stewart, E. E. M., Valsecchi, M., & Schütz, A. C. (2020). A review of interactions between peripheral and foveal vision. *Journal of Vision*, *20*(11), 1–35. https://doi.org/10.1167/jov.20.12.2

Talcott, T. N., Kiat, J. E., Luck, S. J., & Gaspelin, N. (2023). Is covert attention necessary for programming accurate saccades? Evidence from saccade-locked event-related potentials. *Attention, Perception, and Psychophysics*. https://doi.org/10.3758/s13414-023-02775-5

Tatler, B. W., Brockmole, J. R., & Carpenter, R. H. S. (2017). Latest: A model of saccadic decisions in space and time. *Psychological Review*, *124*(3), 267–300. https://doi.org/10.1037/rev0000054

Tatler, B. W., Hayhoe, M. M., Land, M. F., & Ballard, D. H. (2011). Eye guidance in natural vision: Reinterpreting salience. *Journal of Vision*, *11*(5), 1–23. https://doi.org/10.1167/11.5.1

Tatler, B. W., & Vincent, B. T. (2008). Systematic tendencies in scene viewing. *Journal of Eye Movement Research*, *2*(2), 1–18. https://doi.org/10.16910/jemr.2.2.5

Thielen, J., Bosch, S. E., van Leeuwen, T. M., van Gerven, M. A. J., & van Lier, R. (2019). Evidence for confounding eye movements under attempted fixation and active viewing in cognitive neuroscience. *Scientific Reports*, *9*(1), 1–8. https://doi.org/10.1038/s41598-019-54018-z

Turatto, M., Valsecchi, M., Tamè, L., & Betta, E. (2007). Microsaccades distinguish between global and local visual processing. *NeuroReport*, *18*(10), 1015–1018. https://doi.org/10.1097/WNR.0b013e32815b615b

Vanderwal, T., Eilbott, J., & Castellanos, F. X. (2019). Movies in the magnet: Naturalistic paradigms in developmental functional neuroimaging. *Developmental Cognitive Neuroscience*, *36*(May 2018), 100600. https://doi.org/10.1016/j.dcn.2018.10.004

Vinken, K., Prince, J. S., Konkle, T., & Livingstone, M. S. (2023). The neural code for "face cells" is not face-specific. *Science Advances*, *9*(35), 1–14. https://doi.org/10.1126/SCIADV.ADG1736

Visconti di Oleggio Castello, M., Chauhan, V., Jiahui, G., & Gobbini, M. I. (2020). An fMRI dataset in response to "The Grand Budapest Hotel", a socially-rich, naturalistic movie. *Scientific Data*, *7*(1), 1–9. https://doi.org/10.1038/s41597-020-00735-4

Wallisch, P., & Whritner, J. A. (2017). Strikingly low agreement in the appraisal of motion pictures. *Projections (New York)*, *11*(1), 102–120. https://doi.org/10.3167/proj.2017.110107

Wang, L., Baumgartner, F., Kaule, F. R., Hanke, M., & Pollmann, S. (2019). Individual face- and house-related eye movement patterns distinctively activate FFA and PPA. *Nature Communications*, *10*(1), 1–16. https://doi.org/10.1038/s41467-019-13541-3

Wilmott, J. P., & Michel, M. M. (2021). Transsaccadic integration of visual information is predictive, attention-based, and spatially precise. *Journal of Vision*, *21*(8). https://doi.org/https://doi.org/10.1167/jov.21.8.14

Xiao, W., Sharma, S., Kreiman, G., & Livingstone, M. S. (2024). Feature-selective responses in macaque visual cortex follow eye movements during natural vision. *Nature Neuroscience*, 1–10. https://doi.org/10.1038/s41593-024-01631-5

Xu, J., Jiang, M., Wang, S., Kankanhalli, M. S., & Zhao, Q. (2014). Predicting human gaze beyond pixels. *Journal of Vision*, *14*(1), 1–20. https://doi.org/10.1167/14.1.28

Xu-Wilson, M., Zee, D. S., & Shadmehr, R. (2009a). The intrinsic value of visual information affects saccade velocities. *Molecular and Cellular Biochemistry*, *23*(1), 1–7. https://doi.org/10.1007/s00221-009-1879-1.The

Xu-Wilson, M., Zee, D. S., & Shadmehr, R. (2009b). The intrinsic value of visual information affects saccade velocities. *Experimental Brain Research*, *196*(4), 475–481. https://doi.org/10.1007/s00221-009-1879-1

Yoon, T., Jaleel, A., Ahmed, A. A., & Shadmehr, R. (2020). Saccade vigor and the subjective economic value of visual stimuli. *Journal of Neurophysiology*, *123*(6), 2161–2172. https://doi.org/10.1152/jn.00700.2019

Zhang, Y., Kim, J. H., Brang, D., & Liu, Z. (2021). Naturalistic stimuli: A paradigm for multiscale functional characterization of the human brain. *Current Opinion in Biomedical Engineering*, *19*, 1–17. https://doi.org/10.1016/j.cobme.2021.100298

7 Publications

7.1 Study 1

Borovska, P., & de Haas, B. (2023). Faces in scenes attract rapid saccades. *Journal of Vision*, *23*(8), 1–15. https://doi.org/10.1167/jov.23.8.11

Faces in scenes attract rapid saccades

Petra Borovska Experimental Psychology, Justus Liebig University, Giessen, Germany

Benjamin de Haas Experimental Psychology, Justus Liebig University, Giessen, Germany

During natural vision, the human visual system has to process upcoming eye movements in parallel to currently fixated stimuli. Saccades targeting isolated faces are known to have lower latency and higher velocity, but it is unclear how this generalizes to the natural cycle of saccades and fixations during free-viewing of complex scenes. To which degree can the visual system process high-level features of extrafoveal stimuli when they are embedded in visual clutter and compete with concurrent foveal input? Here, we investigated how free-viewing dynamics vary as a function of an upcoming fixation target while controlling for various low-level factors. We found strong evidence that face- versus inanimate object–directed saccades are preceded by shorter fixations and have higher peak velocity. Interestingly, the boundary conditions for these two effects are dissociated. The effect on fixation duration was limited to face saccades, which were small and followed the trajectory of the preceding one, early in a trial. This is reminiscent of a recently proposed model of perisaccadic retinotopic shifts of attention. The effect on saccadic velocity, however, extended to very large saccades and increased with trial duration. These findings suggest that multiple, independent mechanisms interact to process high-level features of extrafoveal targets and modulate the dynamics of natural vision.

Introduction

A crucial question in sensory neuroscience is how foveated visual systems combine the processing of upcoming eye movements with that of currently fixated stimuli to manage the alternating flow of fixations and saccades. A vast literature on transsaccadic integration shows that features of an upcoming target can be processed before a saccade is initiated (Herwig & Schneider, 2014; Osterbrink & Herwig, 2021; Wilmott & Michel, 2021). In tasks presenting isolated stimuli, face-directed saccades show lower latency (Broda & de Haas, 2022a; Crouzet, Kirchner, & Thorpe, 2010) and higher velocity (Xu-Wilson, Zee, & Shadmehr, 2009) than those directed to control inanimate objects. However, it is unclear to which degree this translates to gaze dynamics during the natural cycle of saccades and fixations during free-viewing. In natural scenes, the upcoming target typically is embedded in visual clutter, and the programming of a saccade occurs in parallel to the processing of the currently foveated stimulus. Do faces affect gaze dynamics under these conditions in a similar way?

An effect of faces on peak velocity

It has long been thought that peak velocity forms a stereotypical relationship with saccade amplitude, which is insensitive to changes in stimulus properties (Xu-Wilson et al., 2009). This relationship is referred to as "main sequence" (Bahill, Clark, & Stark, 1975): Peak velocity increases linearly with amplitude, up to a saturation point (Rigas, Komogortsev, & Shadmehr, 2016). Later studies have used saccadic choice paradigms and isolated stimuli to show that this saturation point, as well as the steepness of the linear fit, can differ between observers (Reppert, Lempert, Glimcher, & Shadmehr, 2015) and crucially also be increased for faces as targets (Kauffmann et al., 2019; Xu-Wilson et al., 2009). The study by Xu-Wilson et al. (2009) has shown that saccades to locations expected to show isolated face stimuli, compared to isolated inanimate objects or random pixel noise, had higher velocities and shorter duration, although the effect was relatively small (5.48 dva/s higher for faces on average). A recent study by Yoon, Geary, Ahmed, and Shadmehr (2018) suggests that isolated faces can be understood as items with high reward value, provoking increased vigor (i.e., effort to reach them quickly).
Saccades toward a suddenly appearing stimulus in a saccadic choice task are, however, mostly reactive and may thus differ substantially from voluntary saccade generation during free-viewing (Gremmler & Lappe, 2017; Xu-Wilson et al., 2009). Moreover, natural scene viewing is marked by visual clutter and the concurrent

Citation: Borovska, P., & de Haas, B. (2023). Faces in scenes attract rapid saccades. *Journal of Vision*, *23*(8):11, 1–15, https://doi.org/10.1167/jov.23.8.11.

https://doi.org/10.1167/jov.23.8.11 Received December 20, 2022; published August 8, 2023 ISSN 1534-7362

processing of foveal and extrafoveal input. It is unclear whether the velocity advantage for face-directed saccades generalizes to such more natural free-viewing conditions.

An effect of faces on preceding fixation duration

A predictive model of saccade behavior during free-viewing of naturalistic scenes can be improved by including a shift of attention to the upcoming target location already during the preceding fixation (Schwetlick, Rothkegel, Trukenbrod, & Engbert, 2020). According to this model, this kind of preview contributes to the decision on how long to stay at the currently fixated location. Fixation duration has indeed been shown to be modulated by low-level properties of the upcoming target such as contrast and saturation (Einhäuser, Atzert, & Nuthmann, 2020). However, as discussed in a variety of studies focusing on currently foveated stimuli (Henderson & Pierce, 2008; Kümmerer & Bethge, 2021; Kümmerer, Wallis, Gatys, & Bethge, 2017; Xu, Jiang, Wang, Kankanhalli, & Zhao, 2014), such low-level properties do not fully account for gaze dynamics and high-level, semantic features can improve model performance. One of the most salient types of semantic targets in natural scenes are faces. A number of eye-tracking studies have shown that faces are preferentially targeted (Coutrot & Guyader, 2014; Foulsham, Cheng, Tracy, Henrich, & Kingstone, 2010) and fixated longer than other types of inanimate objects during free-viewing of natural scenes (Guo, Mahmoodi, Robertson, & Young, 2006). Whether faces as targets also modulate the duration of the *preceding* fixation during free-viewing is not entirely clear.

As mentioned, lower saccadic latencies for faces have been found in saccadic choice tasks (Broda & de Haas, 2022b; Crouzet et al., 2010), in which isolated stimuli suddenly appear in opposite hemifields and participants have to saccade to a predefined semantic target category. These tasks use a "gap design" in which the preceding fixation dot disappears just before the onset of target and distractor to minimize its effect on latency and have documented "ultrarapid" saccades with latencies as low as 100 ms toward faces. This is in stark contrast to natural viewing conditions, in which the currently fixated part of a scene and the upcoming target have to be processed in parallel and targets are embedded in the scene and thus visual clutter (Nuthmann, 2017).

Few studies have investigated to which degree lower saccadic latencies in choice tasks generalize to shorter preceding fixations during free-viewing. Cerf, Paxon Frady, and Koch (2009) found that the very first saccade directed toward a scene had lower latency when it was directed toward faces or text rather than cell phones. Similarly, Martin, Davis, Riesenhuber, and Thorpe (2018) recently found that "ultrarapid" saccades generalize to faces superimposed on a scene background. Most important, Mackay, Cerf, and Koch (2012) found that the first few saccades on a complex scene (following the initial one) could be preceded by short fixations and predicted by a salience model, including an explicit face channel. However, fixation durations during scene viewing are known to be shaped by several oculomotor and low-level factors (Tatler & Vincent, 2008), which were not considered in these previous studies. For example, the angle and amplitude of an incoming saccade can predict the magnitude of the following (outgoing) saccade and in turn the duration of the intermittent fixation (Schwetlick et al., 2020; Tatler & Vincent, 2008; Tatler, Brockmole, & Carpenter, 2017). Specifically, short saccades are likely to be followed by saccades in either a similar or the opposite direction, and a fixation between two saccades with similar direction is likely to be of short duration (Schwetlick et al., 2020). Moreover, target size and low-level saliency features such as local luminance contrast at the current and target locations can impact fixation duration (Dick, Ostendorf, Kraft, & Ploner, 2004; Tatler et al., 2017). As of yet, it is unclear how such oculomotor and low-level factors may interact with or confound the effect of faces on fixation durations during free-viewing.

Taken together, previous findings suggest that faces as targets provoke low-latency, high-velocity saccades. However, it is unclear to which degree these effects generalize to free-viewing, especially when controlling for other known factors of oculomotor dynamics. Here, we used a large data set of more than 100 observers free-viewing hundreds of complex scenes, containing close to 50,000 relevant saccadic events (around 40,000 inanimate object–directed and 7,000 face-directed saccades complying with stringent selection criteria). This allowed us to test whether human viewing dynamics are modulated by semantic properties of the upcoming saccade target during the natural cycle of fixations and saccades, taking into account a range of low-level factors known to modulate gaze dynamics. Specifically, we compared the peak velocity of face-directed versus inanimate object–directed saccades and the duration of preceding (inanimate object–directed) fixations. We hypothesized that saccades targeting faces (1) have higher peak velocity and (2) are preceded by shorter fixation durations. The size of our data set allowed us to control for a range of potential confounds and modulators occurring under natural viewing conditions that have been reported to affect saccade latency and/or velocity: saccadic amplitude of the incoming and target saccades (cf. Figure 1; Tatler & Vincent, 2008; Xu-Wilson et al., 2009), trial time (Nuthmann, 2017; Tatler et al., 2017), relative angle of incoming and outgoing saccades (Schwetlick et al., 2020), target size (Dick

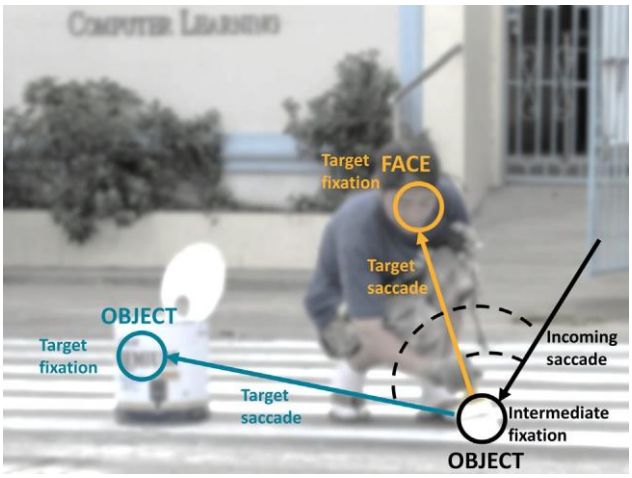

Figure 1. A sequence of incoming saccade, intermediate fixation, target saccade, and target fixation overlaid on an example image. We identified face-related and inanimate object–related saccade as target saccades that landed either on an inner face region (orange example) or an inanimate object (cyan example). Our dependent variables were the peak velocity of the target saccade and the duration of the preceding intermediate fixation. Independent control variables included the amplitude and peak velocity of the incoming saccade and the angle between incoming saccade and target saccade (dashed lines). Note the example image shown has been blurred for illustrative purposes.

et al., 2004; Guadron, van Opstal, & Goossens, 2022), and low-level salience at the preceding fixation and target locations (Einhäuser et al., 2020; Harel, Koch, & Perona, 2007; Nuthmann, 2017). We controlled for these predictors because we expected that they matter for peak velocity and/or preceding fixation duration based on previous literature. Specifically, previous findings suggest that peak velocity increases with target saccade amplitude (i.e., the main sequence; Bahill et al., 1975), peak velocity decreases across trial time (Unema, Pannasch, Joos, & Velichkovsky, 2005), fixation duration increases with the angle between incoming and target saccades (Schwetlick et al., 2020), and fixation duration increases across trial time (Unema et al., 2005). However, we had no strong expectations how these predictors would interact with the effect of faces and consider the corresponding analyses exploratory.

To foreshadow our results, we found clear evidence that face-directed saccades have higher peak velocities and are preceded by shorter fixation durations. Interestingly, the effect of shorter preceding fixation durations is limited to face-directed saccades with relatively low amplitudes, following inanimate object–directed saccades of similar direction and occurring early in a trial. This may point to an interaction between face-channel and saccade-related retinotopic shifts of attention (Mackay et al., 2012; Schwetlick et al., 2020). At the same time, the effect on saccadic velocity generalizes to large saccades and increases over trial time, suggesting that multiple, dissociable mechanisms process high-level features outside the fovea to modulate gaze dynamics.

Materials and methods

Participants

We reanalyzed an existing data set of 103 participants free-viewing 700 complex scenes. The fixations of these participants were previously analyzed and published (Linka & de Haas, 2020). Here, we extracted and analyzed their saccades. Subjects were recruited at Leibniz Institute of Psychology Trier using the PsychLab offline service. We excluded two subjects from the analysis due to missing data files, leaving a sample of $N = 101$ ($M_{age} = 25.21$; $SD = 5.54$; 8 left-handed; 70 females). All participants had normal or corrected-to-normal vision. The study was approved by the local ethics committee and all participants gave informed consent before the experiment. For details, see Linka and de Haas (2020).

Apparatus

Participants placed their heads in a chin and forehead rest and viewed stimuli at a distance of ∼64 cm at 29.7 × 22.3 degrees visual angle. The experiment was controlled via Psychtoolbox (Kleiner et al., 2007) and MATLAB (MathWorks, Natick, MA, USA). Gaze data were acquired using an EyeLink 1000 Plus eye tracker (SR Research, Ottawa, Canada) at a frequency of 2 kHz.

Stimuli and procedure

We used annotated stimuli from the Object and Semantic Images and Eye-tracking (OSIE) data set (Xu et al., 2014). The OSIE contains a total of 700 complex everyday scenes and corresponding pixel masks for 5,551 *visual objects* (we refer to *visual objects* as a superordinate category including both face and inanimate object) with binary labeling for 12 semantic attributes (e.g., *Faces*, *Text*, *Touched*). For details, see Xu et al. (2014). Additionally, we used OSIEplus masks and labels (Broda & de Haas, 2022a), which refine the pixel masks for persons into nine categories (e.g., *Inner faces*, *Heads*, *Eyes*). For details,

see Broda and de Haas (2022b). Participants freely viewed all 700 images in seven blocks of 100 images each. Each block was preceded by a calibration. Before each image presentation, a self-paced fixation disk appeared, followed by a display of the image for 3 s. All images were presented in the same order across participants.

Data and availability

Anonymized data and MATLAB code to reproduce the presented findings are freely available at https://osf.io/vj985/.

Analysis

Preprocessing

Saccades and fixations were extracted by using the SR Research saccade detection algorithm and parser with default values of a minimum velocity of 30°/s and a minimum acceleration of 8,000°/s^2. Gaze coordinates were mapped to image coordinates and removed if they fell outside of the image borders. To exclude fixations and saccades initiated before image onset, fixations and saccades with an onset time < 100 ms trial time were disregarded, which amounted to 9% of saccades and fixations exclusion on average (Linka & de Haas, 2020; SR Research, 2022). Additionally, fixations with a duration under 100 ms were excluded (SR Research, 2022). This led to an exclusion of 5% of fixations on average. To prevent erroneous gaze estimation during lid occlusion caused by a blink, saccades occurring 100 ms before or after a blink were also discarded (i.e., 5% of fixations on average were removed). We further removed potential corrective saccades (i.e., 0.3% of saccades on average were removed). Corrective saccades were defined as saccades that were smaller than 30% of the preceding saccade and had an angle deviation less than 20 degrees (same-directed) or more than 160 degrees (opposite-directed) to that previous saccade. We also disregarded saccades and fixations with a duration > 1,000 ms (Nuthmann, 2017) or peak velocity > 1,000 deg/s. That led to an exclusion of 0.2% of fixations and saccades on average.

Event detection

We identified events of interest for each trial and each participant as intermediate fixations that were preceded and followed by saccades, which we refer to as incoming and target saccade, respectively. This process necessarily excluded the last fixations of the trial and the first saccade of the trial. To label fixations and saccades as falling on a given *visual object*, we used the OSIE pixel masks (see above). We used the additional OSIEplus pixel masks (Broda & de Haas, 2022a) to identify fixations on the *inner face* region of a depicted person (thus excluding, e.g., fixations on the back of the head). A fixation was assigned the label(s) of a given pixel mask if a radius of ∼0.5 degrees visual angle around the nominal fixation center overlapped with the mask (i.e., the approximate area of foveation). We additionally required saccades to have start and landing points on different *visual objects* (Linka & de Haas, 2021). Intermediate fixations had to be on inanimate objects and target fixations on inanimate objects (inanimate object–directed saccades) or the inner region of a human face (face-directed saccades). We also excluded all animal-related saccades (see Supplementary Table S10 for details on frequency of faces, animals, and inanimate objects). This resulted in 6,809 valid face-directed and 42,072 valid inanimate object–directed target saccades across participants and images. Note there could be multiple valid event series for a given observer and image. Figure 1 shows a valid event series, consisting of incoming saccade–intermediate fixation–inanimate object or face-directed target saccade–target fixation.

Parameters of interest

To test the potential effect of semantic target category (face vs. inanimate object) on saccade latency and velocity, we tested whether the duration of intermediate fixations (in ms) and the peak velocity of target saccades (in deg/s) varied as a function of target. To test potential interactions and control for potential confounds, we considered several additional independent variables that have been reported to affect saccade latency and/or velocity: amplitude of the target saccade in degrees visual angle (dva) (Tatler & Vincent, 2008; Xu-Wilson et al., 2009), absolute angle (Schwetlick et al., 2020) of the target saccade relative to the incoming saccade in degrees (deg) (Figure 1; with 0 denoting a continuation and 180 a reversal), onset time of the target saccade relative to image onset in ms (i.e., time in trial; Nuthmann, 2017; Tatler et al., 2017), target size (i.e., area of the corresponding pixel mask, expressed as percentage of image area; Dick et al., 2004; Guadron et al., 2022) (see Supplementary Figure S6 for details on size distribution), and graph-based visual saliency (GBVS) (Einhäuser et al., 2020; Harel et al., 2007; Nuthmann, 2017) at the intermediate and target fixation locations (i.e., sum of pixel saliency values in a radius of ∼0.5 dva around the fixation center).

Statistical analysis

To compare the peak velocity of target saccades landing on faces and inanimate objects, as well as the duration of preceding intermediate fixations, we extracted all relevant events for participants (N = 101)

and trials ($N = 700$). Statistical tests were conducted in MATLAB R2020b (MathWorks) using the *ttest*, *anovan*, and *fitlme* functions.

We used separate linear mixed-effects models to test for an effect of face versus inanimate object (semantic target category) on target saccade peak velocity and intermediate fixation duration. We used dummy coding for semantic target category, with faces coded as 1 and the reference category of inanimate objects coded as 0. In addition to semantic target category, we included seven further predictors to control for potential confounds (see above): (1) target saccade amplitude, (2) incoming saccade amplitude, (3) size of target stimuli, (4) time from onset of the trial, (5) angle of the target to incoming amplitude, (6) GBVS of intermediate fixation, and (7) GBVS of target fixation. All continuous predictor variables were *z*-scored. The dependent variable peak velocity was *z*-scored and the dependent variable fixation duration was *z*-scored and log-transformed due to the right-skewness of the underlying distribution. We used three random factors in both models: *subject* (101 levels), *image* (591 levels), and *visual object* (2,857 levels). The *images* were crossed with *subjects*, and the *visual objects* were nested in *images* (see Supplementary Table S1 for details). We estimated both an intercept and a slope for *subject* and *image* but not for *visual object* as it was either a face or an inanimate object. We selected the best-fitting model specification based on differences in Akaike's information criterion (AIC), considering both main and random effects. To do this, we iteratively removed one fixed predictor at a time from the model and compared all candidate models to the one with minimal AIC:

$$f::_i = AIC_i - AIC_{min}$$

where AIC_{min} is the AIC of the model with the lowest AIC among all candidate models, AIC_i is the AIC of the *i*th other candidate model, and $f::_i$ designates the difference between their AICs (Burnham & Anderson, 2002). Both models showed the lowest AIC for the full model with all predictors included. If available, we selected the most simple model performing on par with the full model according to AIC (i.e., $f::_i < 2$; Burnham & Anderson, 2002). This was the case only for the fixation duration model, including a random by-*subject* intercept and slope, random by-*image* intercept, and random by-*visual object* intercept (Supplementary Table S2 & Supplementary Table S3). For peak velocity, we selected the full model, including a random by-*subject* slope and intercept, random by-*image* slope and intercept, and random by-*visual object* intercept (Supplementary Table S2 & Supplementary Table S3).

We also estimated covariance parameters for random effects (Supplementary Table S3) and conducted model diagnostics, visually inspecting residual plots (Supplementary Figure S1 & Supplementary Figure S2). The full table of linear mixed-effects model (LMM) results including AIC comparisons is reported in Supplementary Tables S2 to S7.

Furthermore, we ran seven two-way analyses of variance (ANOVAs) for each of the dependent variables of interest (target saccade peak velocity and intermediate fixation duration). Each ANOVA tested a potential interaction effect between semantic target category (inanimate object or face) and one control variable. Specifically, the control variables tested in the seven ANOVAs were (1) the angle between incoming and target saccade, distributed across 15 bins of 12 degrees each; (2) target saccade amplitude, distributed across 14 bins of 1 dva each; (3) target saccade onset time, distributed across 13 bins of 200 ms; (4) target size, distributed across 10 bins of 10% each; (5) GBVS of intermediate fixation, distributed across 10 bins of 40 arbitrary salience units (a.u.) each; (6) GBVS of target fixation, distributed across 40 a.u. each; and (7) incoming saccade amplitude, distributed across 14 bins of 1 dva each. We expected three of these predictors to be of particular importance: the angle between incoming and target saccade, the target saccade amplitude, and target saccade onset time. We show the corresponding ANOVA results in the main text. The full list of ANOVA results is reported in the Supplementary Table S8. For each ANOVA, we ran separate post hoc paired *t*-tests. The significance level of these *t*-tests was determined at a family-wise error rate of $\alpha = 0.05$ using the Holm–Bonferroni method to correct for multiple testing (asterisks in plots denote significance surviving this correction).

As a post hoc control analysis, we used linear mixed-effects models to test for an effect of time from trial onset and the angle between incoming and target saccades on the amplitude of target saccades. This model also included semantic target category as a control predictor for specific effects of faces and inanimate objects as targets on saccadic amplitude. We controlled for these effects to further explore an initial finding of slower saccades toward the end of the trial (Supplementary Figure S5 & Supplementary Table S9).

Finally, we conducted a control analysis to test whether the effect of faster saccades toward faces and shorter preceding fixation durations is driven by animacy and extends to human bodies (Yun, Peng, Samaras, Zelinsky, & Berg, 2013). We contrasted saccades landing on human bodies (without faces) versus inanimate objects. This resulted in 5,594 valid body-related and 42,072 inanimate object–related target saccades. Again, we used separate linear mixed-effects models to test for an effect of body versus inanimate object on target saccade peak velocity and intermediate fixation durations. Model specifications were identical to the main analysis contrasting faces and inanimate objects. We did not find evidence supporting an effect of animacy and therefore do not report follow-up ANOVAs and *t*-tests here and instead refer interested readers to our OSF repository (osf.io/vj985/).

Results

Events of interest

We used data from 101 participants free-viewing 700 complex everyday scenes for 3 s each. We identified events of interest for each trial and participant as an intermediate fixation landing on any inanimate object, followed by a saccade targeting either a face or another inanimate object (Figure 1). Pooled across participants and trials, we found 6,809 such events targeting a face and 42,072 targeting an inanimate object. Figure 1 shows a valid event series, consisting of incoming saccade–intermediate fixation–inanimate object or face-directed target saccade–target fixation.

Main hypotheses

We tested two main hypotheses regarding the influence of an upcoming face or inanimate object target on free-viewing gaze behavior: (1) Peak velocity will be higher for a saccade targeting faces versus inanimate objects, and (2) the duration of an intermediate inanimate object fixation will be shorter when the following saccade target is a face versus inanimate object. To test simple main effects of semantic target category (face vs. inanimate object), we used linear mixed-effects models for each measure. To control for potential confounds (see Introduction), we included seven additional predictors: (1) target saccade amplitude, (2) incoming saccade amplitude, (3) size of target stimuli, (4) time from onset of the trial, (5) angle of the target to incoming amplitude, (6) GBVS at the intermediate fixation, and (7) GBVS at the target fixation. To test potential modulatory effects of these low-level factors, we additionally ran two-way ANOVAs, testing potential interactions between semantic target category and one control variable at a time.

Saccade peak velocity

The average peak velocity of saccades targeting a face ($N = 6{,}809$, $M = 369.48$, $SD = 104.28$, $SE = 1.26$) was indeed higher than that of saccades targeting inanimate objects ($N = 42{,}072$, $M = 328.99$, $SD = 112.24$, $SE = 0.54$).

Linear mixed-effect model of semantic target category and control predictors

To test the statistical significance of this effect and control for potential confounds, we implemented a linear mixed-effect model (for full results, see Supplementary Table S2 & Supplementary Table S3). This confirmed strong evidence for an effect of semantic target category ($b = 0.07$, $SE = 0.02$, $t(48{,}872) = 4.33$, $p < 0.001$) (Figure 2d), indicating a higher peak velocity for saccades targeting faces versus inanimate objects, even when other relevant predictors were held constant. The model also confirmed the expected strong effect of target saccade amplitude ($b = 0.74$, $SE = 0.003$, $t(48{,}872) = 238.72$, $p < 0.001$), that is, the main sequence (Bahill et al., 1975). Expressed in standardized weights, the effect of semantic target category amounted to about 10% of that observed for target saccade amplitudes. Additional significant but smaller effects on peak velocity included trial time of saccade onset ($b = 0.02$, $SE = 0.002$, $t(48{,}872) = 8.93$, $p < 0.001$; indicating saccadic velocity increases as trial time progresses), target size ($b = 0.01$, $SE = 0.005$, $t(48{,}872) = 2.52$, $p < 0.05$; indicating saccadic velocity increased as the size of the target increased), low-level saliency of the target ($b = 0.02$, $SE = 0.003$, $t(48{,}872) = 5.28$, $p < 0.001$; indicating higher velocity saccades toward targets with higher low-level saliency), and amplitude of the incoming saccade ($b = 0.02$, $SE = 0.002$, $t(48{,}872) = 7.43$, $p < 0.001$; indicating higher velocity target saccades for larger incoming saccades). Finally, the effects of low-level saliency at the intermediate location ($b = 0.004$, $SE = 0.002$, $t(48{,}872) = 1.55$, $p = 0.12$) and angle between incoming and target saccade ($b = 0.002$, $SE = 0.002$, $t(48{,}872) = 0.98$, $p = 0.32$) were small and not statistically significant. For details, see Supplementary Table S2 and Supplementary Table S3.

Targeted two-way ANOVAs

We additionally ran two-way ANOVAs to test potential interactions between semantic target category and the remaining predictors. The first of these models tested the simple main effects of target category, $F(1, 48{,}880) = 777.4$, $p < 0.05$, $\eta^2 = 0.015$, and deviation of the target saccade angle from that of the incoming saccade, $F(14, 48{,}880) = 38.35$, $p < 0.05$, $\eta^2 = 0.011$. Unlike in the LMM results, the simple main effect of angle was significant. Holm–Bonferroni corrected post hoc paired *t*-tests showed faster saccades toward faces versus inanimate objects across all relative angles (Figure 2a). However, there also was a significant interaction, $F(14, 48{,}880) = 4.48$, $p < 0.05$, $\eta^2 = 0.001$, indicating a higher velocity advantage for face-targeting saccades with a similar angle to the incoming one.

The second two-way ANOVA tested and confirmed significant simple main effects of target category, $F(1, 45{,}048) = 53.27$, $p < 0.05$, $\eta^2 = 0.001$, and target saccade amplitude, $F(13, 45{,}048) = 1{,}703.03$, $p < 0.05$, $\eta^2 = 0.33$, but no significant interaction between the two, $F(13, 45{,}048) = 0.77$, $p = 0.68$, $\eta^2 = 0.0002$. As shown in Figure 2b and Supplementary Table S8,

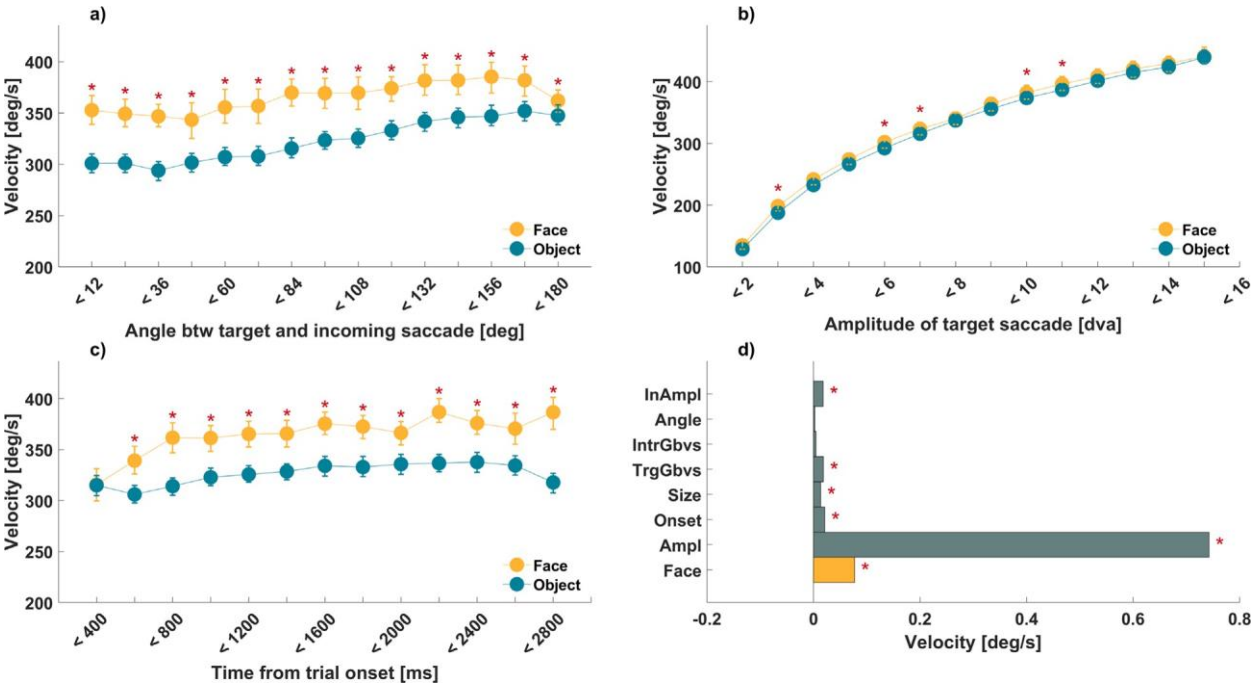

Figure 2. Peak velocity. (**a**–**c**) Peak velocity of target saccades landing on faces and inanimate objects in cyan and yellow, as shown in the inset. Red asterisks mark Bonferroni-corrected significance of paired *t*-test and error bars represent bootstrapped 95% confidence interval (1,000 resamples). (**a**) Peak velocity as a function of absolute deviation of saccade angles between target and incoming saccade. The increase in angle represents an increase from same-directed saccades (<12 degrees) to opposite-directed saccades (<180 degrees). (**b**) Peak velocity as a function of target saccade amplitude (in dva), showing the main sequence. (**c**) Peak velocity as a function of time from onset (ms) within a trial. (**d**) Standardized, fitted predictor weights of a linear mixed-effects model of peak velocity with simple main effects of semantic target category (Face; shown in yellow bar), target amplitude (Ampl), time from trial onset (Onset), size of target stimuli (Size), low-level salience at target (TrgGbvs) and intermediate (IntrGbvs) fixation, absolute deviation of saccade angle between incoming and target saccade (Angle), and amplitude (InAmpl) of the incoming saccade. Red asterisks mark statistically significant beta coefficients. Note that all continuous variables were *z*-scored and thus the corresponding beta values indicate effects in standard deviation units.

Holm–Bonferroni corrected post hoc *t*-tests indicated significantly faster saccades toward faces versus inanimate objects for most amplitude bins, up to 12 dva (with an average advantage of 7.18 deg/s).

The third two-way ANOVA tested and confirmed significant simple main effects of target category, $F(1, 48{,}875) = 696.88$, $p < 0.05$, $\eta^2 = 0.01$, and trial onset time, $F(12, 48{,}875) = 21.37$, $p < 0.05$, $\eta^2 = 0.005$. Figure 2c shows significant Holm–Bonferroni corrected post hoc *t*-tests for almost all onset bins, except very early saccades under 400 ms. A significant interaction, $F(12, 48{,}875) = 5.75$, $p < 0.05$, $\eta^2 = 0.001$, pointed to a general increase of the velocity advantage for face-directed saccades over trial time.

Remaining predictors

For completeness, we also ran additional ANOVAs for all remaining predictors (cf. Supplementary Results, Table S8). We found significant interactions between target category and each of the following factors: the target size ($F(9, 48{,}070) = 24.01$, $p < 0.05$, $\eta^2 = 0.004$; indicating a higher velocity advantage for smaller faces), low-level saliency of the intermediate fixation location ($F(9, 13{,}727) = 2.39$, $p < 0.05$, $\eta^2 = 0.001$; indicating a more pronounced velocity advantage for faces when low-level saliency at the intermediate location was high), and amplitude of the incoming saccade ($F(13, 41{,}496) = 2.41$, $p < 0.05$, $\eta^2 = 0.0008$; indicating a higher velocity advantage for faces for higher amplitudes of the incoming saccade).

Interim summary, Part 1

Taken together, these results showed higher velocities for face- compared to inanimate object–directed saccades. This effect was substantial (about 10% of that of the main sequence) and robust to controlling for a range of other factors it interacted with. These interactions point to a velocity effect of faces for saccades of all amplitudes, increasing over trial time, being largest for saccades continuing a large incoming saccade in a straight line and when low-level salience at the intermediate fixation location is high.

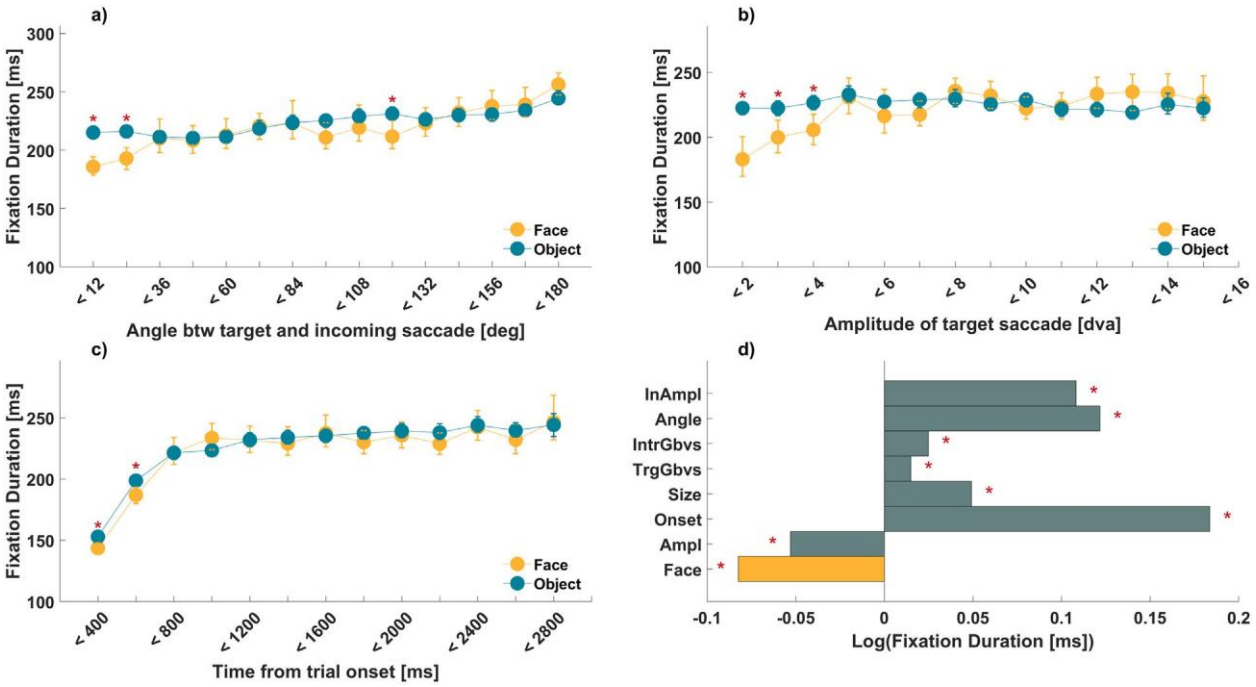

Figure 3. Measures of fixation duration. (**a**–**c**) Intermediate fixation duration (ms) followed by saccade landing on faces and inanimate objects in yellow and cyan, as shown in the inset. Red asterisks mark Bonferroni-corrected significance of paired *t*-test and error bars represent bootstrapped 95% confidence interval (1,000 resamples). (**a**) Intermediate fixation duration as a function of absolute deviation of saccade angles between target and incoming saccade. The increase in angle represents an increase from same-directed saccades (<12 degrees) to opposite-directed saccades (<180 degrees). (**b**) Intermediate fixation duration as a function of target saccade amplitude (in dva). (**c**) Intermediate fixation duration as a function of time from onset (ms) within a trial. (**d**) Standardized, fitted predictor weights of a linear mixed-effects model of intermediate fixation duration with simple main effects of semantic target category (Face; shown in yellow bar), target amplitude (Ampl), time from trial onset (Onset), size of target stimuli (Size), low-level salience of target (TrgGbvs) and intermediate (IntrGbvs) fixation, absolute deviation of saccade angle between incoming and target saccade (Angle), and amplitude of the incoming saccade (InAmpl). Red asterisks mark statistically significant beta coefficients. Note that all continuous variables were *z*-scored and thus the corresponding beta values indicate effects in standard deviation units.

Fixation duration

We compared the fixation duration at intermediate fixation locations in milliseconds when the following saccade landed on faces (N = 6,809, M = 224.38, SD = 106.4 , SE = 1.29) versus inanimate objects (N = 42,072, M = 224.70, SD = 101.77, SE = 0.49). The simple means of fixation durations for both types of fixations were not statistically different.

Linear mixed-effect model of semantic target category and control predictors

We implemented a similar linear mixed-effect model as for peak velocity to test the statistical significance of the effect of face targets on intermediate fixation durations when other predictors were held constant. We observed a significant, negative effect of semantic target category (b = –0.08, SE = 0.02, t(48,872) = –4.09, p < 0.001; Figure 3d), indicating shorter intermediate fixation for face-directed target saccades. A range of further predictors had significant and sometimes strong effects on the duration of intermediate fixations: time from trial onset (b = 0.18, SE = 0.004, t(48,872) = 40.96, p < 0.001; indicating longer intermediate fixation durations as trial time progress), absolute angle between target and incoming saccade (b = 0.12, SE = 0.005, t(48,872) = 26.09, p < 0.001; indicating longer intermediate fixation durations preceding saccades that reversed direction), amplitude of the incoming saccade (b = 0.11, SE = 0.005, t(48,872) = 23.38, p < 0.001; indicating longer intermediate fixation durations for larger incoming saccades), amplitude of target saccade (b = –0.05, SE = 0.005, t(48,872) = –9.53, p < 0.001; indicating longer intermediate fixation durations for smaller target saccades), size of the target (b = 0.04, SE = 0.007, t(48,872) = 6.76, p < 0.001; indicating longer intermediate fixation durations when the following saccade landed on a larger target), low-level salience at the intermediate fixation location (b = 0.02, SE = 0.005, t(48,440) = 4.62, p < 0.001; indicating longer intermediate fixation durations for higher low-level saliency of the currently fixated inanimate

object), and low-level salience of the target ($b = 0.01$, $SE = 0.005$, $t(48{,}872) = 4.62$, $p < 0.001$; indicating longer intermediate fixation durations for higher low-level saliency of the target). For details, see Supplementary Table S2 and Supplementary Table S3.

Targeted two-way ANOVAs

Similar as for peak velocity, we ran additional two-way ANOVAs to test potential interactions between semantic target category and other predictors in modulating intermediate fixation duration. The first, two-way ANOVA revealed significant simple main effects of semantic target category, $F(1, 48{,}880) = 10.12$, $p < 0.05$, $\eta^2 = 0.0002$, and absolute deviation angle between the incoming and target saccades, $F(14, 48{,}880) = 33.85$, $p < 0.05$, $\eta^2 = 0.009$. A significant interaction of semantic category and angle, $F(14, 48{,}880) = 4.73$, $p < 0.05$, $\eta^2 = 0.001$, indicated the shortest fixation durations preceded face-targeting saccades in the same direction as the incoming saccade, but this effect of faces was diminished or even reversed for saccades going in the opposite direction. This was confirmed by post hoc Holm–Bonferroni corrected paired *t*-test showing shorter fixation durations for faces in same-directed saccades, $t(97) = -7.1$, $p < 0.05$, $t(93) = -4.69$, $p < 0.05$, and for saccades in almost opposite directions (angle of 120 degrees), $t(96) = -3.3$, $p < 0.05$ (Figure 3e).

A second two-way ANOVA showed a significant simple main effect of semantic target category, $F(1, 45{,}048) = 4.83$, $p < 0.05$, $\eta^2 = 0.0001$, and target saccade amplitude, $F(13, 45{,}048) = 3.81$, $p < 0.05$, $\eta^2 = 0.001$. A significant interaction, $F(13, 45{,}048) = 4.23$, $p < 0.05$, $\eta^2 = 0.001$, indicated that shorter fixation durations precede face-directed saccades of small amplitudes. This was confirmed by post hoc Holm–Bonferroni corrected paired *t*-tests, which only showed significant face effects for the shortest amplitudes for 2 dva, $t(64) = -4.09$, $p < 0.05$; 3 dva, $t(94) = -3.18$, $p < 0.05$; and 4 dva, $t(91) = -3.71$, $p < 0.05$ (Figure 3f).

A third two-way ANOVA revealed significant simple main effects of semantic category, $F(1, 48{,}875) = 5.91$, $p < 0.05$, $\eta^2 = 0.0001$, and onset time, $F(12, 48{,}875) = 91.7$, $p < 0.05$, $\eta^2 = 0.02$. The interaction between these factors missed statistical significance, $F(12, 48{,}875) = 1.73$, $p = 0.053$, $\eta^2 = 0.0004$. Nevertheless, shorter fixation duration preceding face-directed saccades seemed limited to early trial times up to 600 ms, as indicated by Holm–Bonferroni corrected paired *t*-test, $t(95) = -4.35$, $p < 0.05$, $t(91) = -3.3$, $p < 0.05$ (cf. Figure 3g).

Remaining predictors

For completeness, we also ran additional ANOVAs for all remaining predictors (cf. Supplementary Results, Table S8). Although some of the interactions regarding the effect of faces as targets were significant, these effects appeared unsystematic.

Interim summary, Part 2

Taken together, when controlling for a range of potential confounds, intermediate fixation durations depended on the following saccade target. Fixations preceding face-directed saccades were shorter than those preceding saccades directed to inanimate objects. This effect was small to moderate compared to other factors, such as trial time, and interacted with two other predictors: Intermediate fixations are shortest before small face-directed saccades at an angle continuing the incoming saccade. Additionally, the effect appears limited to the first 600 ms of a trial.

Fixation duration and peak velocity toward bodies

To probe whether rapid saccades to faces are due to a general animacy effect (Yun et al., 2013), we repeated the main analyses for saccades targeting bodies instead of faces. We implemented linear mixed-effect models to test for an effect of bodies versus inanimate objects as targets on the peak velocity of target saccades and on the duration of preceding intermediate fixations.

We observed a nonsignificant, negative effect of target category on peak velocity ($b = -0.008$, $SE = 0.01$, $t(47{,}657) = -0.58$, $p = 0.55$; Figure 4a), indicating no evidence for body-targeting saccades to be faster than saccades targeting inanimate objects. As in the main analysis, there was a range of further significant predictors: the target saccade amplitude ($b = 0.75$ $SE = 0.003$, $t(47{,}657) = 245.73$, $p < 0.001$; indicating increasing saccade speed with amplitude), time from trial onset ($b = 0.01$, $SE = 0.002$, $t(47{,}657) = 6.91$, $p < 0.001$; indicating saccadic velocity increases as trial time progresses), low-level salience of the target ($b = 0.01$, $SE = 0.003$, $t(47{,}657) = 5.18$, $p < 0.001$; indicating higher velocity saccades toward targets with higher low-level saliency), and amplitude of the incoming saccade ($b = 0.01$, $SE = 0.002$, $t(47{,}657) = 7.08$, $p < 0.001$; indicating higher velocity target saccades for larger incoming saccades).

Similarly, a linear mixed-effects model showed a reversed effect on preceding intermediate fixation durations, with fixations preceding body-directed saccades lasting longer than saccades preceding inanimate object–directed saccades ($b = 0.04$, $SE = 0.01$, $t(47{,}657) = 2.08$, $p < 0.05$). Further significant predictors showed a similar profile to those in the main analysis: time from trial onset ($b = 0.18$, $SE = 0.004$, $t(47{,}657) = 40.63$, $p < 0.001$; indicating longer intermediate fixation durations as trial time progress), absolute angle between target and incoming saccade ($b = 0.11$, $SE = 0.004$, $t(47{,}657) = 24.63$,

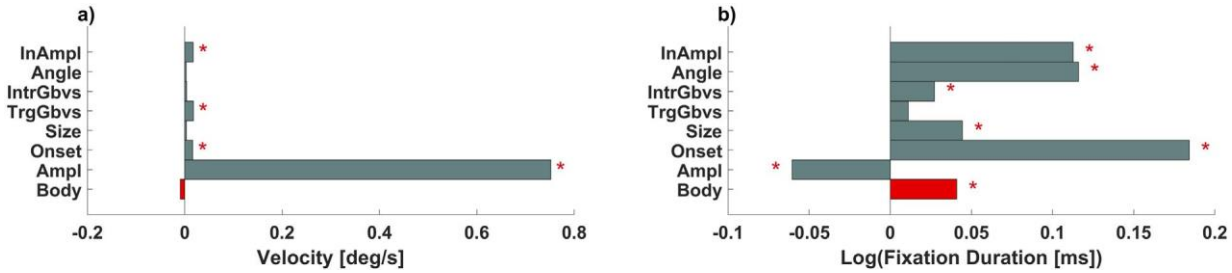

Figure 4. Linear mixed-effects models fitting standardized velocity and log fixation duration. Standardized, fitted predictor weights of a linear mixed-effects model of peak velocity (**a**) and intermediate fixation duration (**b**) with simple main effects of target category (Body; shown in red bar), target amplitude (Ampl), time from trial onset (Onset), size of target stimuli (Size), low-level salience of target (TrgGbvs) and intermediate (IntrGbvs) fixation, absolute deviation of saccade angle between incoming and target saccade (Angle), and amplitude of the incoming saccade (InAmpl). Red asterisks mark statistically significant beta coefficients. Note that all continuous variables were *z*-scored and thus the corresponding beta values indicate effects in standard deviation units.

$p < 0.001$; indicating longer intermediate fixation durations preceding saccades that reversed direction), amplitude of the incoming saccade ($b = 0.11$, $SE = 0.004$, $t(47{,}657) = 24.04$, $p < 0.001$; indicating longer intermediate fixation durations for larger incoming saccades), amplitude of target saccade ($b = -0.06$, $SE = 0.005$, $t(47{,}657) = -10.92$, $p < 0.001$; indicating longer intermediate fixation durations for smaller target saccades), size of the target ($b = 0.04$, $SE = 0.007$, $t(47{,}657) = 6.31$, $p < 0.001$; indicating longer intermediate fixation durations when the following saccade landed on a larger target), and low-level salience at the intermediate fixation location ($b = 0.02$, $SE = 0.005$, $t(47{,}657) = 5.03$, $p < 0.001$; indicating longer intermediate fixation durations for higher low-level saliency of the currently fixated inanimate object).

Interim summary, Part 3

We found no evidence for rapid saccades toward bodies. Body-directed saccades were not significantly different from saccades toward inanimate objects with respect to peak velocity (Figure 4a). Regarding saccadic latency, fixations preceding body-directed saccades lasted *longer* than those preceding saccades toward inanimate objects (Figure 4b).

Discussion

Shorter fixation and higher peak velocity as evidence of extrafoveal processing of high-level features

In the present study, we investigated whether high-level properties of extrafoveal *visual objects* in a complex scene can modulate free-viewing dynamics *before* they are fixated. We found strong evidence that face- versus inanimate object–directed saccades are preceded by shorter fixations and have higher peak velocity. These results are in line with previous findings on the latency and velocity advantage for saccades directed to isolated faces (Broda, Haddad, & de Haas, 2022; Crouzet et al., 2010; Kauffmann, Khazaz, Peyrin, & Guyader, 2021; Reppert et al., 2015; Xu-Wilson et al., 2009; Yoon, Jaleel, Ahmed, & Shadmehr, 2020) and show it extends to free-viewing complex scenes, which is marked by visual clutter and the concurrent processing of foveal and extrafoveal input. The concurrent processing of high-level features at both currently foveated and target locations is matching findings from the transsaccadic literature showing that features of the upcoming target can be processed before the saccade is initiated (Herwig & Schneider, 2014; Osterbrink & Herwig, 2021; Wilmott & Michel, 2021). It also matches the notion that peripheral vision is enhanced by foveal feedback, aiding object recognition (Stewart, Valsecchi, & Schütz, 2020). The rich data set we used allowed us to control for a range of potential confounds and moderators, revealing that the effect of faces on free-viewing dynamics is modulated by target eccentricity, the trajectory of consecutive saccades, and the time from trial onset. Taken together, our results provide strong evidence for the extrafoveal processing of high-level features in natural vision and reveal related moderators that point to potential underlying mechanisms.

Faces in scenes

Human gaze behavior is systematic, and much research has been devoted to predicting where humans look in a scene when and for how long. Two major approaches (Henderson, 2011) focus on (1) features of the scene (e.g.,

Itti & Koch, 2000) or (2) top-down control (e.g., Yarbus, 1965). More recent efforts are trying to combine both and emphasize high-level features of scenes (e.g., Einhäuser, Spain, & Perona, 2008; Xu et al., 2014) and faces in particular (Cerf, Harel, Einhäuser, & Koch, 2008; Kümmerer, Wallis, & Bethge, 2016; Xu et al., 2014). In laboratory paradigms, a number of studies have shown that faces are preferentially targeted (Coutrot & Guyader, 2014; Foulsham et al., 2010) and longer fixated (Guo et al., 2006), and saccades toward them tend to be faster (Xu-Wilson et al., 2009). Faces are deemed high-value targets by several studies (Xu-Wilson et al., 2009; Yoon et al., 2018), and adding a face channel to low-level saliency models significantly improves gaze prediction (Cerf et al., 2008). Our results show that faces, but not bodies or inanimate objects, attract rapid saccades during scene viewing. This corroborates the special role of faces for human gaze behavior (see Results; Figure 4).

Studies of occipitotemporal face processing find a strong central visual field bias (e.g., Levy, Hasson, Avidan, Hendler, & Malach, 2001). This is interesting in light of our findings, which show a strong central bias for the effect of faces on saccadic latency (i.e., preceding fixation durations) but not for that on saccadic velocity, which generalized to the periphery (also see below).

The effect of faces on saccadic velocity

Face-directed saccades had higher peak velocity compared to inanimate object–directed saccades, and this effect of target held even when controlling for various other factors known to modulate saccade velocity. The main factor determining peak velocity is amplitude, resulting in the main sequence relationship (Bahill et al., 1975; Reppert et al., 2015). Interestingly, the effect of faces on peak velocity was constant throughout the amplitude spectrum, even for large saccades.

Velocity also increased with trial time and did so more strongly for face- versus inanimate object–directed saccades. This is in contrast to the general scene-viewing tendency of shorter saccade amplitudes and thus slower saccades toward the end of viewing time (Unema et al., 2005). This apparent discrepancy can be explained by the fact that we limited our analysis to saccade events moving from one *visual object* to another, and many of the small, slow saccades toward the end of a trial inspect successive details within a given *visual object* (focal vs. ambient mode; Nuthmann, 2017; Trevarthen, 1968). Indeed, our data show that the magnitude of saccades for the types of events we selected *increased* over time (see Supplementary Figure S5 & Supplementary Table S9). A possible explanation is that saccades can move more easily between *visual objects* at a greater distance toward the end of a trial, because the target or close-by regions have been visited previously.

The velocity advantage for face-directed saccades was also somewhat larger when incoming and target saccades followed the same trajectory. Importantly, however, the velocity of face-directed saccades is higher than that of inanimate object–directed saccades, independently of low-level salience. This is in line with the importance of semantic features for fixation locations in complex scenes (Itti & Koch, 2000; Mackay et al., 2012; Xu et al., 2014) and extends their importance to the corresponding saccade dynamics. Previous studies found higher velocities for speeded saccades to isolated faces (Xu-Wilson et al., 2009). This was interpreted to reflect a high intrinsic reward value of faces, as targets associated with reward, such as food (Takikawa, Kawagoe, Itoh, Nakahara, & Hikosaka, 2002) or monetary profit (Chen, Chen, Zhou, & Mustain, 2014), that have been shown to elicit saccades with an increased velocity profile as well. Our results suggest that this effect holds for natural vision too (i.e., the free-viewing of complex scenes), which is marked by visual clutter and the concurrent processing of foveal and extrafoveal input. Free-viewing typically elicits self-paced voluntary as opposed to reactive saccades (Gremmler & Lappe, 2017). Interestingly, the overall peak velocity advantage we observe for faces versus inanimate objects (advantage of 7.18 deg/s) is even larger than that which has been reported in the context of reactive saccades (advantage of 5.48 deg/s; Xu-Wilson et al., 2009). A modulation of viewing dynamics during natural vision appears adaptive. It could help with the time-critical prioritization of conspecifics in visual clutter or of targets in a foraging situation, such as searching for fruit in a canopy.

The effect of faces on preceding fixation durations

Saccades targeting faces versus inanimate objects were preceded by shorter fixations when controlling for potentially confounding predictors. This is reminiscent of the very low latencies observed for saccades directed to isolated face stimuli (Broda et al., 2022; Crouzet et al., 2010; Martin et al., 2018). Our results show this effect extends to free-viewing, where extrafoveal targets are processed concurrently with currently foveated targets. This seems remarkable, given saccadic choice paradigms typically use a gap design, in order to avoid any concurrent foveal input, including that of a fixation dot.

Importantly, the duration effect we observed here is limited to fixations preceding small saccades following a trajectory similar to the preceding one. This

matches the hypothesis of a perisaccadic attentional spotlight shifting in retinotopic coordinates when the saccade is executed (Schwetlick et al., 2020). Schwetlick et al. (2020) recently provided modeling evidence for the decoupling of covert attention and current fixation position in target selection, followed by a brief retinotopic shift of attention in the direction of the saccade, until it is realigned with the current fixation position. Our current results suggest that the resulting pull along the saccadic trajectory is especially pronounced if the shifted postsaccadic window of attention falls on a parafoveal face.

Interestingly, the effect of upcoming face targets on fixation duration seemed limited to saccades occurring within the first 600 ms of a trial, which is in line with previous findings (Mackay et al., 2012) and reminiscent of the notion of ambient versus focal processing (Nuthmann, 2017; Trevarthen, 1968). This co-occurred with a general increase of fixation duration with time from trial onset, which is a well-established finding (Tatler et al., 2017; Unema et al., 2005).

Our findings also suggest that an upcoming face target can only shorten intermediate fixation durations when the face is closer than 4 dva. This may be related to the perisaccadic attentional shifts discussed above, because saccades of a similar direction to the preceding one tend to be small (Schwetlick et al., 2020). It may also point to the involvement of face-selective neurons with a strong central bias in their visual field coverage (see above and below). An important caveat is that most of the faces in the stimulus set we used were rather small (most inner face regions extend < 3 dva). Target size was a positive predictor of preceding fixation durations in our data set, but we cannot rule out that very large faces may shorten saccadic latencies at higher eccentricities. This could be investigated in future studies sampling face eccentricities and sizes more systematically and comprehensively.

Taken together, the effect of face targets on preceding fixation durations appeared more limited than that on peak velocity. Fixation durations were only modulated at the beginning of a trial and for nearby faces, whereas the effect on peak velocity was observed even for the largest saccades and *increased* across trial duration. This suggests that the threshold for an extrafoveal target to modulate the preceding fixation duration may be higher than for modulating saccadic velocity. For instance, the memory of a face at a peripheral location may be sufficient to elicit a saccade with higher velocity, whereas a shortening of the preceding fixation may require the direct parafoveal registering of a face that has not been fixated before. Although speculative at this point, this may also be reflected in the underlying biological mechanisms. The modulation of preceding fixation durations may require the activation of face-sensitive neurons in the ventral stream, which have a strong central bias in their visual field coverage (Gomez, Natu, Jeska, Barnett, & Grill-Spector, 2018; Issa & Dicarlo, 2012; Kay, Weiner, & Grill-Spector, 2015), whereas the modulation of peak velocity may rest on a different (possibly subcortical) face channel with wider visual field coverage.

Future research and limitations

In terms of future research, even bigger data sets would allow to examine possible effects of semantic features beyond the face–inanimate object distinction, for both intermediate and target *visual objects*.

Although we constrained intermediate fixations to inanimate objects, we cannot rule out the possibility that the effects we observed are modulated by semantic features at this and the target location. For example, text (Cerf et al., 2009; Mackay et al., 2012) or food may be capable of eliciting fast, low-latency saccades as well, and the saccadic "pull" of such features most likely interacts with that of the intermediate fixation target. Future studies could also consider intermediate face fixations and compare face-to-face versus face-to-inanimate object saccades. This may require targeted stimuli and controls, given that scenes with multiple faces tend to come with compositional biases (e.g., faces appearing at the same height). Finally, given the evidence for strong individual traits in gaze behavior (Bargary et al., 2017; Broda & de Haas, 2022a, 2022b; de Haas, Iakovidis, Schwarzkopf, & Gegenfurtner, 2019; Linka, Broda, Alsheimer, de Haas, & Ramon, 2022; Linka & de Haas, 2020; Rigas et al., 2016; Yoon et al., 2020), even larger data sets may allow individual estimates of the effects we found here. Testing the interindividual covariance of latency and velocity effects could provide valuable evidence regarding a core hypothesis suggested by our results: The effect of faces on saccadic latency and velocity may rest on separate mechanisms.

Conclusion

In summary, we found evidence that faces in complex scenes elicit rapid saccades. Face-directed saccades have higher peak velocity across the amplitude spectrum. This effect is substantial (about 10% of that of the main sequence) and increases across the duration of a trial and for saccades following the trajectory of the preceding saccade. It may reflect mechanisms utilizing memory of previously fixated face locations and/or processes with a wide visual field coverage. Face-directed saccades are also preceded by shorter fixation durations. However, this effect is limited

to small saccades early in a trial, which follow the trajectory of the preceding one. This may reflect the perisaccadic shift of an attentional window and face processing mechanisms with a strong parafoveal bias. Thus, the dynamics of natural vision appear to be modulated by several interacting mechanisms, allowing the processing of high-level features outside the fovea.

Keywords: face, extrafoveal, complex scenes, fixation duration, saccade peak velocity

Acknowledgments

The authors thank Maximilian D. Broda for his OSIEplus pixel masks and Marcel Linka and Diana Kollenda for discussion and corrections to this manuscript.

Supported by European Research Council Starting Grant 852885 INDIVISUAL; BdH was further supported by Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) Project No. 222641018–SFB/TRR 135 TP C9 and "The Adaptive Mind," funded by the Excellence Program of the Hessian Ministry of Higher Education, Science, Research and Art.

Commercial relationships: none. Corresponding author: Petra Borovska.
Email: petra.borovska@psychol.uni-giessen.de.
Address: Department of Psychology, Justus Liebig University, Giessen, Germany.

References

Bahill, A. T., Clark, M. R., & Stark, L. (1975). The main sequence, a tool for studying human eye movements. *Mathematical Biosciences, 24*(3–4), 191–204, https://doi.org/10.1016/0025-5564(75)90075-9.

Bargary, G., Bosten, J. M., Goodbourn, P. T., Lawrance-Owen, A. J., Hogg, R. E., & Mollon, J. D. (2017). Individual differences in human eye movements: An oculomotor signature? *Vision Research, 141*, 157–169, https://doi.org/10.1016/j.visres.2017.03.001.

Burnham, K. P., & Anderson, D. R. (2002) *Model selection and inference: A practical information-theoretic approach.* (2nd ed. p. 488). New York: Springer, http://dx.doi.org/10.1007/b97636.

Broda, M. D., & de Haas, B. (2022a). Individual differences in looking at persons in scenes. *Journal of Vision, 22*(12), 9, https://doi.org/10.1167/jov.0.0.08318.

Broda, M. D., & de Haas, B. (2022b). Individual fixation tendencies in person viewing generalize from images to videos. *i-Perception, 12*(2), 1–10, https://doi.org/10.1177/20416695211009552.

Broda, M. D., Haddad, T., & de Haas, B. (2022). Quick, eyes! Isolated upper face halves but not artificial features elicit rapid saccades. *PsyArxiv*, https://doi.org/10.31234/osf.io/24gsj.

Cerf, M., Harel, J., Einhäuser, W., & Koch, C. (2008). Predicting human gaze using low-level saliency combined with face detection. *Advances in Neural Information Processing Systems, 20*, 241–248.

Cerf, M., Paxon Frady, E., & Koch, C. (2009). Faces and text attract gaze independent of the task: Experimental data and computer model. *Journal of Vision, 9*(12), 1–15, https://doi.org/10.1167/9.12.1.

Chen, L. L., Chen, Y. M., Zhou, W., & Mustain, W. D. (2014). Monetary reward speeds up voluntary saccades. *Frontiers in Integrative Neuroscience, 8*, 48, https://doi.org/10.3389/fnint.2014.00048.

Coutrot, A., & Guyader, N. (2014). How saliency, faces, and sound influence gaze in dynamic social scenes. *Journal of Vision, 14*(8), 1–17, https://doi.org/10.1167/14.8.5.

Crouzet, M., Kirchner, H., & Thorpe, S. J. (2010). Fast saccades toward faces: Face detection in just 100 ms. *Journal of Vision, 10*(4), 16, https://doi.org/10.1167/10.4.16.

de Haas, B., Iakovidis, A. L., Schwarzkopf, D. S., & Gegenfurtner, K. R. (2019). Individual differences in visual salience vary along semantic dimensions. *Proceedings of the National Academy of Sciences of the United States of America, 116*(24), 11687–11692, https://doi.org/10.1073/pnas.1820553116.

Dick, S., Ostendorf, F., Kraft, A., & Ploner, C. J. (2004). Saccades to spatially extended targets: The role of eccentricity. *NeuroReport, 15*(3), 453–456, https://doi.org/10.1097/00001756-200403010-00014.

Einhäuser, W., Atzert, C., & Nuthmann, A. (2020). Fixation durations in natural scene viewing are guided by peripheral scene content. *Journal of Vision, 20*(4), 1–15, https://doi.org/10.1167/jov.20.4.15.

Einhäuser, W., Spain, M., & Perona, P. (2008). Objects predict fixations better than early saliency. *Journal of Vision, 8*(14), 18, https://doi.org/10.1167/8.14.18.

Foulsham, T., Cheng, T. J., Tracy, J. L., Henrich, J., & Kingstone, A. (2010). Gaze allocation in a dynamic situation: Effects of social status and speaking. *Cognition, 117*(3), 319–331, https://doi.org/10.1016/j.cognition.2010.09.003.

Gomez, J., Natu, V., Jeska, B., Barnett, M., & Grill-Spector, K. (2018). Development differentially

sculpts receptive fields across early and high-level human visual cortex. *Nature Communications, 9*(1), 788, https://doi.org/10.1038/s41467-018-03166-3.

Gremmler, S., & Lappe, M. (2017). Saccadic suppression during voluntary versus reactive saccades. *Journal of Vision, 17*, 1–10, https://doi.org/10.1167/17.8.8.doi.

Guadron, L., van Opstal, A. J., & Goossens, J. (2022). Speed-accuracy tradeoffs influence the main sequence of saccadic eye movements. *Scientific Reports*, *12*(1), 1–14, https://doi.org/10.1038/s41598-022-09029-8.

Guo, K., Mahmoodi, S., Robertson, R. G., & Young, M. P. (2006). Longer fixation duration while viewing face images. *Experimental Brain Research, 171*(1), 91–98, https://doi.org/10.1007/s00221-005-0248-y.

Harel, J., Koch, C., & Perona, P. (2007). Graph-based visual saliency. *Advances in Neural Information Processing Systems, 19*, 545–552, https://doi.org/10.7551/mitpress/7503.003.0073.

Henderson, J. M. (2011). Eye movements and scene perception. In S. P. Liversedge, I. D. Gilchrist, & S. Everling (Eds.), *The Oxford handbook of eye movements* (pp. 593–606). Oxford University Press.

Henderson, J. M., & Pierce, G. L. (2008). Eye movements during scene viewing: Evidence for mixed control of fixation durations. *Psychonomic Bulletin & Review, 15*, 566–573, https://doi.org/10.3758/PBR.15.3.566.

Herwig, A., & Schneider, W. X. (2014). Predicting object features across saccades: Evidence from object recognition and visual search. *Journal of Experimental Psychology: General, 143*(5), 1903–1922, https://doi.org/10.1037/a0036781.

Issa, E. B., & Dicarlo, J. J. (2012). Precedence of the eye region in neural processing of faces. *Journal of Neuroscience*, *32*(47), 16666–16682, https://doi.org/10.1523/JNEUROSCI.2391-12.2012.

Itti, L., & Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research, 40*(10–12), 1489–1506, https://doi.org/10.1016/S0042-6989(99)00163-7.

Kauffmann, L., Khazaz, S., Peyrin, C., & Guyader, N. (2021). Isolated face features are sufficient to elicit ultra-rapid and involuntary orienting responses toward faces. *Journal of Vision, 21*(2), 1–24, https://doi.org/10.1167/jov.21.2.4.

Kauffmann, L., Peyrin, C., Chauvin, A., Entzmann, L., Breuil, C., & Guyader, N. (2019). Face perception influences the programming of eye movements. *Scientific Reports, 9*(1), 1–14, https://doi.org/10.1038/s41598-018-36510-0.

Kay, K. N., Weiner, K. S., & Grill-Spector, K. (2015). Attention reduces spatial uncertainty in human ventral temporal cortex. *Current Biology, 25*(5), 595–600, https://doi.org/10.1016/j.cub.2014.12.050.

Kleiner, M., Brainard, D., Pelli, D., Ingling, A., Murray, R., & Broussard, C. (2007). What's new in psychtoolbox-3. *Perception, 36*(14), 1–16.

Kümmerer, M., & Bethge, M. (2021). State-of-the-art in human scanpath prediction. *arXiv*:2102.12239, https://doi.org/10.48550/arXiv.2102.12239.

Kümmerer, M., Wallis, T. S. A., & Bethge, M. (2016). DeepGaze II: Reading fixations from deep features trained on object recognition. ArXiv:1610.01563, https://doi.org/10.48550/arXiv.1610.01563.

Kümmerer, M., Wallis, T. S. A., Gatys, L. A., & Bethge, M. (2017). Understanding low- and high-level contributions to fixation prediction. *2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy* (pp. 4799–4808), https://doi.org/10.1109/ICCV.2017.513.

Levy, I., Hasson, U., Avidan, G., Hendler, T., & Malach, R. (2001). Center-periphery organization of human object areas. *Nature Neuroscience, 4*(5), 533–539, https://doi.org/10.1038/87490.

Linka, M., Broda, M. D., Alsheimer, T., de Haas, B., & Ramon, M. (2022). Characteristic fixation biases in Super-Recognizers. *Journal of Vision, 22*(8), 17, https://doi.org/10.1167/jov.22.8.17.

Linka, M., & de Haas, B. (2020). OSIEshort: A small stimulus set can reliably estimate individual differences in semantic salience. *Journal of Vision, 20*(9), 1–9, https://doi.org/10.1167/JOV.20.9.13.

Linka, M., & de Haas, B. (2021). Detection, inspection and re-inspection: A functional approach to gaze behavior towards complex scenes. *Journal of Vision, 21*(9), 1971, https://doi.org/10.1167/jov.21.9.1971.

Mackay, M., Cerf, M., & Koch, C. (2012). Evidence for two distinct mechanisms directing gaze in natural scenes. *Journal of Vision, 12*(4), 9, https://doi.org/10.1167/12.4.9.

Martin, J. G., Davis, C. E., Riesenhuber, M., & Thorpe, S. J. (2018, November). Zapping 500 faces in less than 100 seconds: Evidence for extremely fast and sustained continuous visual search. *Scientific Reports*, pp. 1–12, https://doi.org/10.1038/s41598-018-30245-8.

Nuthmann, A. (2017). Fixation durations in scene viewing: Modeling the effects of local image features, oculomotor parameters, and task. *Psychonomic Bulletin and Review*, *24*(2), 370–392, https://doi.org/10.3758/ s13423-016-1124-4.

Osterbrink, C., & Herwig, A. (2021). Prediction of complex stimuli across saccades. *Journal of Vision, 21*(2), 1–15, https://doi.org/10.1167/jov.21.2.10.

Reppert, T. R., Lempert, K. M., Glimcher, P. W., & Shadmehr, R. (2015). Modulation of saccade vigor during value–based decision making. *Journal of Neuroscience*, *35*(46), 15369–15378, https://doi.org/10.1523/JNEUROSCI.2621-15.2015.

Rigas, I., Komogortsev, O., & Shadmehr, R. (2016). Biometric recognition via eye movements. *ACM Transactions on Applied Perception*, *13*(2), 1–21, https://doi.org/10.1145/2842614.

Schwetlick, L., Rothkegel, L. O. M., Trukenbrod, H. A., & Engbert, R. (2020). Modeling the effects of perisaccadic attention on gaze statistics during scene viewing. *Communications Biology*, *3*(1), 1–11, https://doi.org/10.1038/s42003-020-01429-8.

SR Research. (2022). *EyeLink 1000 Plus user manual*. SR Research Ltd, Mississauga, Ontario, Canada, https://www.sr-research.com/support/attachment.php?aid=1376.

Stewart, E. E. M., Valsecchi, M., & Schütz, A. C. (2020). A review of interactions between peripheral and foveal vision. *Journal of Vision*, *20*(11), 1–35, https://doi.org/10.1167/jov.20.12.2.

Takikawa, Y., Kawagoe, R., Itoh, H., Nakahara, H., & Hikosaka, O. (2002). Modulation of saccadic eye movements by predicted reward outcome. *Experimental Brain Research*, *142*(2), 284–291, https://doi.org/10.1007/s00221-001-0928-1.

Tatler, B. W., Brockmole, J. R., & Carpenter, R. H. S. (2017). Latest: A model of saccadic decisions in space and time. *Psychological Review*, *124*(3), 267–300, https://doi.org/10.1037/rev0000054.

Tatler, B. W., & Vincent, B. T. (2008). Systematic tendencies in scene viewing. *Journal of Eye Movement Research*, *2*(2), 1–18, https://doi.org/10.16910/jemr.2.2.5.

Trevarthen, C. B. (1968). Two mechanisms of vision in primates. *Psychologische Forschung*, *337*(31), 299–337.

Unema, P. J. A., Pannasch, S., Joos, M., & Velichkovsky, B. M. (2005). Time course of information processing during scene perception: The relationship between saccade amplitude and fixation duration. *Visual Cognition*, *12*(3), 473–494, https://doi.org/10.1080/13506280444000409.

Wilmott, J. P., & Michel, M. M. (2021). Transsaccadic integration of visual information is predictive, attention-based, and spatially precise. *Journal of Vision*, *21*(8), 14, https://doi.org/10.1167/jov.21.8.14.

Xu, J., Jiang, M., Wang, S., Kankanhalli, M. S., & Zhao, Q. (2014). Predicting human gaze beyond pixels. *Journal of Vision*, *14*(1), 1–20, https://doi.org/10.1167/14.1.28.

Xu-Wilson, M., Zee, D. S., & Shadmehr, R. (2009). The intrinsic value of visual information affects saccade velocities. *Molecular and Cellular Biochemistry*, *23*(1), 1–7, https://doi.org/10.1007/s00221-009-1879-1.

Yarbus, A. L. (1965). *Role of eye movements in the visual process*. Moscow, Russia: Nauka.

Yoon, T., Geary, R. B., Ahmed, A. A., & Shadmehr, R. (2018). Control of movement vigor and decision making during foraging. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(44), E10476–E10485, https://doi.org/10.1073/pnas.1812979115.

Yoon, T., Jaleel, A., Ahmed, A. A., & Shadmehr, R. (2020). Saccade vigor and the subjective economic value of visual stimuli. *Journal of Neurophysiology*, *123*(6), 2161–2172, https://doi.org/10.1152/jn.00700.2019.

Yun, K., Peng, Y., Samaras, D., Zelinsky, G. J., & Berg, T. L. (2013). Exploring the role of gaze behavior and object detection in scene understanding. *Frontiers in Psychology*, *4*, Article 917, https://doi.org/10.3389/fpsyg.2013.00917.

7.1.1 Supplement: Study 1

Supplementary Information

1 Mixed-effects Models

1.1 Model Structure

	Name	Nr Levels
Fixed Effects		
	Face	2
	Ampl	continuous
	Onset	continuous
	Size	continuous
	TrgGbvs	continuous
	IntrGbvs	continuous
	Angle	continuous
	InAmpl	continuous
Random Effects		
	Subj	101
	ImgNr (nested in Subj)	591
	ObjNr (nested in ImgNr)	2857

Table S1. Linear mixed-effects models structure for peak velocity and fixation duration. The table shows the structure of models including eight fixed effects terms: semantic category (Face), target amplitude (Ampl), time from trial onset (Onset), size of target stimuli (Size), low-level salience at target fixation (TrgGbvs), low-level salience at intermediate fixation (IntrGbvs), angle of target amplitude to incoming amplitude (Angle) and the amplitude of incoming saccade (InAmpl). And three random effects terms. subject (Subj) with 101 levels, image (ImgNr) with 591 levels (out of 700 images), and visual object (ObjNr) with 2857 levels (out of 5551 objects). We used dummy coding, with faces and inanimate objects coded as 1 and 0, respectively.

1.2 Simple Main Effects

	Velocity			Fixation Duration		
Fixed effects	Estimate	*SE*	t	Estimate	SE	t
(Intercept)	-0.021	0.035	-0.600 n.s.	0.036	0.022	1.610 n.s.
Face	0.077	0.018	4.330*	-0.081	0.021	-3.940*
Ampl	0.742	0.003	238.720*	-0.053	0.006	-9.620*
Onset	0.021	0.002	8.930*	0.184	0.004	40.970*
Size	0.014	0.005	2.520*	0.048	0.007	6.550*
TargetGbvs	0.019	0.004	5.290*	0.015	0.006	2.500*
IntrGbvs	0.005	0.003	1.550 n.s.	0.025	0.005	4.640*
Angle	0.002	0.002	0.990 n.s.	0.121	0.005	26.020*
IncomingAmpl	0.018	0.002	7.430*	0.108	0.005	23.380*
Log-likelihood	-34948.700			-66191.600		
Deviance	69897.400			132383.100		
AIC	69931.400			132417.100		
BIC	70081.000			132566.700		

** p<.05; n.s. = non-significant (p>.05)*

Table S2. **Linear mixed-effects models fitting standardized velocity and log fixation duration.** The table shows all simple main effects of the semantic category (Face), target amplitude (Ampl), time from trial onset (Onset), size of target stimuli (Size), low-level salience at target fixation (TrgGbvs), low-level salience at intermediate fixation (IntrGbvs), angle of target amplitude to incoming amplitude (Angle) and the amplitude of incoming saccade (InAmpl). Asterisks in the column of t-values indicate statistically significant beta coefficients. All continuous predictors were z-scored and fixation duration was additionally log-transformed.

1.3 Random Effects

	Velocity			Fixation Duration		
Random Effects	Term	*SD*	r	Term	*SD*	r
Subj (Levels 101)						
	(Intercept)	0.341		*(Intercept)*	0.194	
	SlopeFace	0.086	0.46	*SlopeFace*	0.059	-0.211
ImgNr (Levels 591)						
	(Intercept)	0.114		*(Intercept)*	0.183	
	SlopeFace	0.172	-0.359			
ObjNr (Levels 2857)						
	(Intercept)	0.167		*(Intercept)*	0.187	
Residual (Error)						
		0.476			0.917	

r denotes correlation between the intercept and the slope of a particular random term

Table S3. **Linear mixed-effects models: Estimates of covariance parameters fitting standardized velocity and log fixation duration.** The table shows random effects estimates for random by-subject (Subj) slope and intercept with 101 levels, random by-image (ImgNr) intercept (and slope only for velocity) with 591 levels, and random by-visual object (ObjNr) intercept with

2857 levels. Standard deviations (SD) are reported for each random term and the residual. Random slope for image number was not part of the winning fixation duration model (see AIC comparisons below; Table S7). Correlation coefficients (r) indicate the relationship between intercepts and slopes for a particular random term.

1.4 Model Diagnostics: Velocity

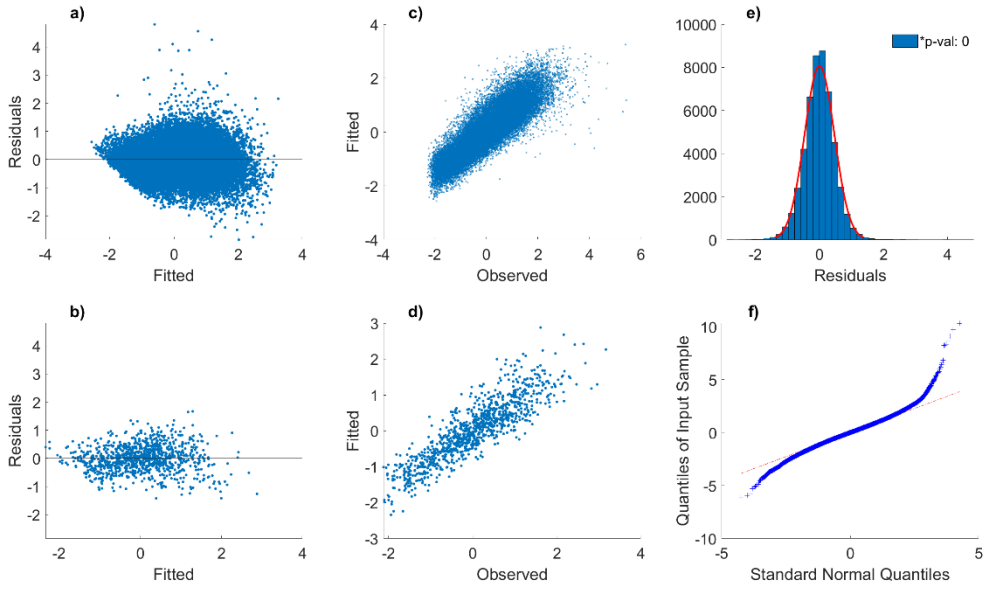

Figure S1. **Diagnostic plots of linear mixed-effects model fitting standardized velocity.** Panel (a) shows residuals as a function of fitted values with the full sample (N = 48,881). To alleviate overplotting, a randomly selected subsample of N = 1000 is shown in panel (b). There is no obvious heteroscedasticity or violation of normality. Panel (c) shows fitted values as a function of observed values, with a clear linear relationship. Similarly, panel (d) shows a subsample of fitted vs. observed values (N = 1000). Panel (e) shows a histogram of residuals, approximating a normal distribution, although a non-parametric test indicated a deviation (Kruskal-Wallis Test, *p-val < .001). Panel (f) shows a Q-Q (quantile-quantile) plot, more revealing of normality violations of residuals at the tails of the distribution. We were not able to achieve a better fit by transforming the data.

$Model_i$ - $Model_{min}$	Δ AIC
Veloc ~ 1 + (Face|Subj) + (Face|ImgNr) + (1|ObjNrRec) - $Model_{min}$	38617.42
Veloc ~ 1 + Face + (Face|Subj) + (Face|ImgNr) + (1|ObjNrRec) - $Model_{min}$	38520.91
Veloc ~ 1 + Face + Ampl + (Face|Subj) + (Face|ImgNr) + (1|ObjNrRec) - $Model_{min}$	157.61
Veloc ~ 1 + Face + Ampl + Onset + (Face|Subj) + (Face|ImgNr) + (1|ObjNrRec) - $Model_{min}$	82.54
Veloc ~ 1 + Face + Ampl + Onset + Size + (Face|Subj) + (Face|ImgNr) + (1|ObjNrRec) - $Model_{min}$	78.56
Veloc ~ 1 + Face + Ampl + Onset + Size + Gbvs + (Face|Subj) + (Face|ImgNr) + (1|ObjNrRec) - $Model_{min}$	53.19
Veloc ~ 1 + Face + Ampl + Onset + Size + Gbvs + CurrGbvs + (Face|Subj) + (Face|ImgNr) + (1|ObjNrRec) - $Model_{min}$	53.55
Veloc ~ 1 + Face + Ampl + Onset + Size + Gbvs + CurrGbvs + Angle + (Face|Subj) + (Face|ImgNr) + (1|ObjNrRec) - $Model_{min}$	53.17
Veloc ~ 1 + Face + Ampl + Onset + Size + Gbvs + CurrGbvs + Angle + PrevAmpl + (Face|Subj) + (Face|ImgNr) + (1|ObjNrRec) - $Model_{min}$	0.00

$Model_{min}$ = Veloc ~ 1 + Face + Ampl + Onset + Size + Gbvs + CurrGbvs + Angle + PrevAmpl + (Face|Subj) + (Face|ImgNr) + (1|ObjNrRec)

Table S4. AIC comparison of fixed effects for linear mixed-effects models fitting standardized velocity. The table shows AIC difference values (Δ) for fixed effects comparing model with lowest AIC ($Model_{min}$) to other candidate models ($Model_i$). Other candidate models were created by always removing one additional main predictor ($Model_i$) and calculating its AIC. As long as the AIC difference (Δ) between minimal and ith model, was no larger than 2, the model with minimal AIC was selected. That was the case for this model of the peak velocity, where the model with minimal AIC was also the most complex model with all respective predictors: Ampl, Onset, Size, Gbvs, CurrGbvs, Angle, and PrevAmpl.

$Model_i$ - $Model_{min}$	Δ AIC
Veloc ~ 1 + Face + Ampl + Onset + Size + Gbvs + CurrGbvs + Angle + PrevAmpl - $Model_{min}$	21912.91
Veloc ~ 1 + Face + Ampl + Onset + Size + Gbvs + CurrGbvs + Angle + PrevAmpl + (1|Subj) - $Model_{min}$	4474.63
Veloc ~ 1 + Face + Ampl + Onset + Size + Gbvs + CurrGbvs + Angle + PrevAmpl + (Face|Subj) - $Model_{min}$	4407.97
Veloc ~ 1 + Face + Ampl + Onset + Size + Gbvs + CurrGbvs + Angle + PrevAmpl + (Face|Subj) + (1|ImgNr) - $Model_{min}$	2116.63
Veloc ~ 1 + Face + Ampl + Onset + Size + Gbvs + CurrGbvs + Angle + PrevAmpl + (Face|Subj) + (1|ImgNr) + (1|ObjNrRec) - $Model_{min}$	30.39
Veloc ~ 1 + Face + Ampl + Onset + Size + Gbvs + CurrGbvs + Angle + PrevAmpl + (Face|Subj) + (Face|ImgNr) + (1|ObjNrRec) - $Model_{min}$	0.00

$Model_{min}$ = *Veloc ~ 1 + Face + Ampl + Onset + Size + Gbvs + CurrGbvs + Angle + PrevAmpl + (Face|Subj) + (Face|ImgNr) + (1|ObjNrRec)*

Table S5. AIC comparison of random effects for linear mixed-effects models fitting standardized velocity. The table shows AIC difference values (Δ) for random effects comparing model with lowest AIC ($Model_{min}$) to other candidate models ($Model_i$). Other candidate models were created by always removing one additional random term ($Model_i$) and calculating its AIC. As long as the AIC difference (Δ) between minimal and ith model, was no larger than 2, the model with minimal AIC was selected. That was the case for this model of the peak velocity, where the model with minimal AIC was also the most complex model with full random structure: (Face|Subj), (Face|ImgNr), and (1|ObjNrRec).

1.5 Model Diagnostics: Fixation Duration

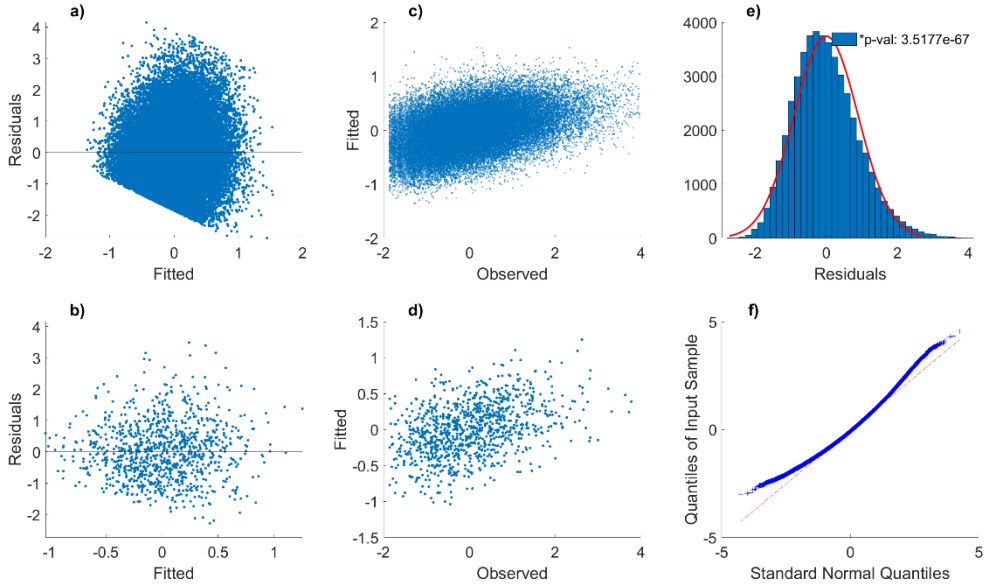

Figure S1. **Diagnostic plots of linear mixed-effects model fitting log fixation duration.** Panel (a) shows residuals as a function of fitted values with the full sample (N = 48,881). Due to a large dataset, randomly selected subsample of N = 1000 is shown in panel (b). There seems to be no obvious pattern suggesting heteroscedasticity or violation of normality, although residuals are cluttered narrowly around the center. Panel (c) shows fitted values as a function of observed values, suggesting a linear relationship. Similarly, panel (d) shows a subsample of fitted vs. observed values (N = 1000). Panel (e) shows a histogram of residuals, approximating a normal distribution, although a non-parametric test indicates some deviation from normality (Kruskal-Wallis Test, *p-val < .001). Panel (f) shows a Q-Q (quantile-quantile) plot revealing minor deviations from normality at the tails.

$Model_i$ - $Model_{min}$	Δ AIC
FixDur ~ 1 + (Face|Subj) + (1|ImgNr) + (1|ObjNrRec) - $Model_{min}$	2821.28
FixDur ~ 1 + Face + (Face|Subj) + (1|ImgNr) + (1|ObjNrRec) - $Model_{min}$	2815.84
FixDur ~ 1 + Face + Ampl + (Face|Subj) + (1|ImgNr) + (1|ObjNrRec) - $Model_{min}$	2810.96
FixDur ~ 1 + Face + Ampl + Onset + (Face|Subj) + (1|ImgNr) + (1|ObjNrRec) - $Model_{min}$	1357.48
FixDur ~ 1 + Face + Ampl + Onset + Size + (Face|Subj) + (1|ImgNr) + (1|ObjNrRec) - $Model_{min}$	1316.25
FixDur ~ 1 + Face + Ampl + Onset + Size + Gbvs + (Face|Subj) + (1|ImgNr) + (1|ObjNrRec) - $Model_{min}$	1303.33
FixDur ~ 1 + Face + Ampl + Onset + Size + Gbvs + CurrGbvs + (Face|Subj) + (1|ImgNr) + (1|ObjNrRec) - $Model_{min}$	1300.94
FixDur ~ 1 + Face + Ampl + Onset + Size + Gbvs + CurrGbvs + Angle + (Face|Subj) + (1|ImgNr) + (1|ObjNrRec) - $Model_{min}$	538.88
FixDur ~ 1 + Face + Ampl + Onset + Size + Gbvs + CurrGbvs + Angle + PrevAmpl + (Face|Subj) + (1|ImgNr) + (1|ObjNrRec) - $Model_{min}$	0.00

$Model_{min}$ = FixDur ~ 1 + Face + Ampl + Onset + Size + Gbvs + CurrGbvs + Angle + PrevAmpl + (Face|Subj) + (1|ImgNr) + (1|ObjNrRec)

Table S6. **AIC comparison of fixed effects for linear mixed-effects models fitting log fixation duration.** The table shows AIC difference values (Δ) for fixed effects comparing model with lowest AIC ($Model_{min}$) to other candidate models ($Model_i$). Other candidate models were created by always removing one additional main predictor ($Model_i$) and calculating its AIC. As long as the AIC difference (Δ) between minimal and ith model, was no larger than 2, the model with minimal AIC was selected. That was the case for this model of the fixation duration, where the model with minimal AIC was also the most complex model with all respective predictors: Ampl, Onset, Size, Gbvs, CurrGbvs, Angle, and PrevAmpl.

$Model_i$ - $Model_{min}$	Δ AIC
FixDur ~ 1 + Face + Ampl + Onset + Size + Gbvs + CurrGbvs + Angle + PrevAmpl - $Model_{min}$	3460.59
FixDur ~ 1 + Face + Ampl + Onset + Size + Gbvs + CurrGbvs + Angle + PrevAmpl + (1|Subj) - $Model_{min}$	1795.25
FixDur ~ 1 + Face + Ampl + Onset + Size + Gbvs + CurrGbvs + Angle + PrevAmpl + (Face|Subj) - $Model_{min}$	1797.24
FixDur ~ 1 + Face + Ampl + Onset + Size + Gbvs + CurrGbvs + Angle + PrevAmpl + (Face|Subj) + (1|ImgNr) - $Model_{min}$	443.44
FixDur ~ 1 + Face + Ampl + Onset + Size + Gbvs + CurrGbvs + Angle + PrevAmpl + (Face|Subj) + (1|ImgNr) + (1|ObjNrRec) - $Model_{min}$	0.77
FixDur ~ 1 + Face + Ampl + Onset + Size + Gbvs + CurrGbvs + Angle + PrevAmpl + (Face|Subj) + (Face|ImgNr) + (1|ObjNrRec) - $Model_{min}$	0.00

$Model_{min}$ = FixDur ~ 1 + Face + Ampl + Onset + Size + Gbvs + CurrGbvs + Angle + PrevAmpl + (Face|Subj) + (Face|ImgNr) + (1|ObjNrRec)

Table S7. AIC comparison of random effects for linear mixed-effects models fitting standardized velocity. The table shows AIC difference values (Δ) for random effects comparing model with lowest AIC ($Model_{min}$) to other candidate models ($Model_i$). Other candidate models were created by always removing one additional random term ($Model_i$) and calculating its AIC. As long as the AIC difference (Δ) between minimal and ith model, was no larger than 2, the model with minimal AIC was selected. This was not the case for this model of fixation duration, as the more simple model had AIC difference lower than 2. Therefore, simpler model of the fixation duration was selected with following random structure: (Face|Subj), (1|ImgNr), and (1|ObjNrRec).

2 Anova

Velocity						Fixation Duration					
Model 1 Angle	*SumSq*	*DF*	*F*	*P*	n^2		*SumSq*	*DF*	*F*	*P*	n^2
Face	9319897.6	1	777.401	p < .05	0.01566	Face	104849.6	1	10.120	p < .05	0.00021
Angle	6437875.4	14	38.357	p < .05	0.01087	Angle	4909407.4	14	33.845	p < .05	0.00961
Face*Angle	752020.2	14	4.481	p < .05	0.00128	Face*Angle	686363.3	14	4.732	p < .05	0.00135
Model 2 Amplitude	*SumSq*	*DF*	*F*	*P*	n^2		*SumSq*	*DF*	*F*	*P*	n^2
Face	237392.1	1	53.271	p < .05	0.00118	Face	50575.2	1	4.825	p < .05	0.00011
Ampl	98660091.4	13	1703.035	p < .05	0.32965	Ampl	519122.6	13	3.810	p < .05	0.00110
Face*Ampl	45022.9	13	0.777	0.685	0.00022	Face*Ampl	576545.2	13	4.231	p < .05	0.00122
Model 3 Onset	*SumSq*	*DF*	*F*	*P*	n^2		*SumSq*	*DF*	*F*	*P*	n^2
Face	8537596.5	1	696.886	p < .05	0.01407	Face	59088.9	1	5.913	p < .05	0.00012
Onset	3142487.1	12	21.376	p < .05	0.00522	Onset	10993901.0	12	91.687	p < .05	0.02203
Face*Onset	845787.8	12	5.753	p < .05	0.00141	Face*Onset	208076.7	12	1.735	0.053	0.00043
Model 4 Size	*SumSq*	*DF*	*F*	*P*	n^2		*SumSq*	*DF*	*F*	*P*	n^2
Face	1217977.3	1	102.777	p < .05	0.00213	Face	2348.6	1	0.226	0.634	0.00000
Size	5955609.1	9	55.839	p < .05	0.01035	Size	865049.4	9	9.252	p < .05	0.00173
Face*Size	2561060.1	9	24.012	p < .05	0.00448	Face*Size	339489.6	9	3.631	p < .05	0.00068
Model 5 Intermediate GBVS	*SumSq*	*DF*	*F*	*P*	n^2		*SumSq*	*DF*	*F*	*P*	n^2
Face	150493.2	1	12.002	p < .05	0.00087	Face	8091.8	1	0.852	0.356	0.00006
IntrGbvs	260722.7	9	2.310	p < .05	0.00151	IntrGbvs	187453.8	9	2.193	p < .05	0.00144
Face*IntrGbvs	270348.6	9	2.396	p < .05	0.00157	Face*IntrGbvs	141075.4	9	1.650	0.095	0.00108
Model 6 Target GBVS	*SumSq*	*DF*	*F*	*P*	n^2		*SumSq*	*DF*	*F*	*P*	n^2
Face	2409579.2	1	180.739	p < .05	0.01423	Face	23738.8	1	2.390	0.122	0.00019
TrgGbvs	1788345.1	9	14.905	p < .05	0.01060	TrgGbvs	258880.6	9	2.897	p < .05	0.00208
Face*TrgGbvs	169578.5	9	1.413	0.176	0.00101	Face*TrgGbvs	136739.7	9	1.530	0.131	0.00110
Model 7 Incoming Amplitude	*SumSq*	*DF*	*F*	*P*	n^2		*SumSq*	*DF*	*F*	*P*	n^2
Face	4744875.7	1	386.560	p < .05	0.00924	Face	5894.9	1	0.603	0.437	0.00001
InAmpl	325753.8	13	2.041	p < .05	0.00064	InAmpl	2456003.1	13	19.325	p < .05	0.00603
Face*InAmpl	385550.8	13	2.416	p < .05	0.00076	Face*InAmpl	225330.6	13	1.773	p < .05	0.00056

Table S8. **Two-way ANOVAs for peak velocity and fixation duration**. Each model tested simple main effects and an interaction of semantic target category with one of the following predictors: absolute deviation of saccade angles between target and incoming saccade (Model 1); target saccade amplitude (Model 2); time from trial onset (Model 3); size of the target stimuli (Model 4)); low-level salience at intermediate fixation (Model 5); low-level salience at target fixation (Model 6); amplitude of the incoming saccade (Model 7).

3 Remaining predictors

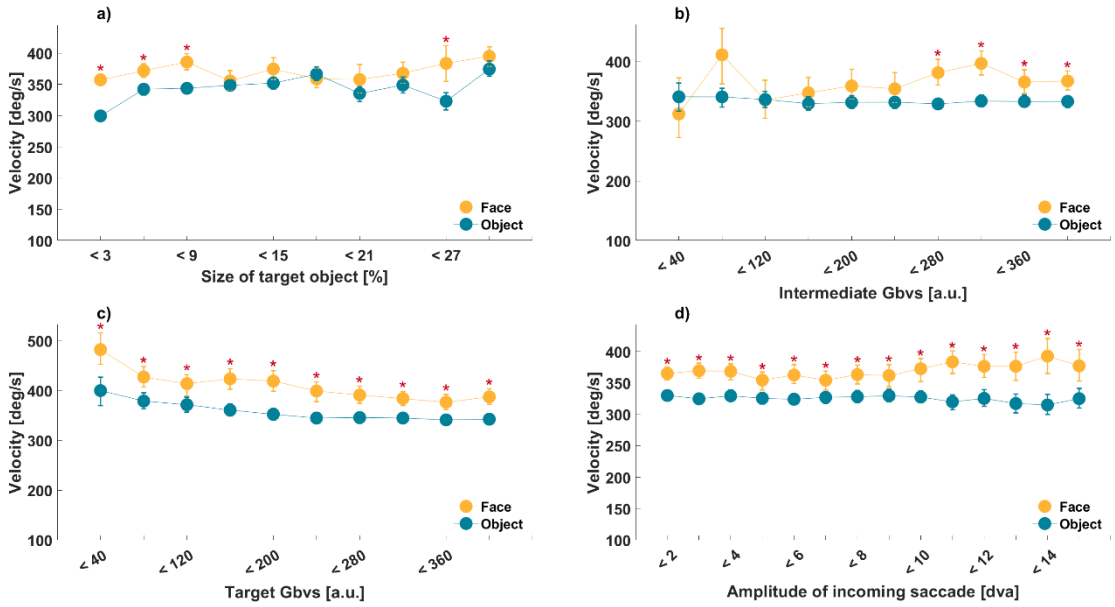

Figure S2. Differences in peak velocity between face- vs. inanimate object-directed saccades. Red asterisks mark Bonferroni corrected significance of paired t-test and error bars represent bootstrapped 95 % confidence interval (1,000 resamples). Panel (a) shows peak velocity as a function of size of target stimuli (Size). Panel (b) shows peak velocity as a function of low-level salience at the intermediate fixation. Panel (c) shows peak velocity as a function of low-level salience at the target fixation. Panel (d) shows peak velocity as a function of amplitude of the incoming saccade. Low-level salience is given in arbitrary units (a.u.); low values indicate low saliency and high values indicate high saliency. Cyan and yellow markers denote data from inanimate object and face-directed saccades as shown in the inset.

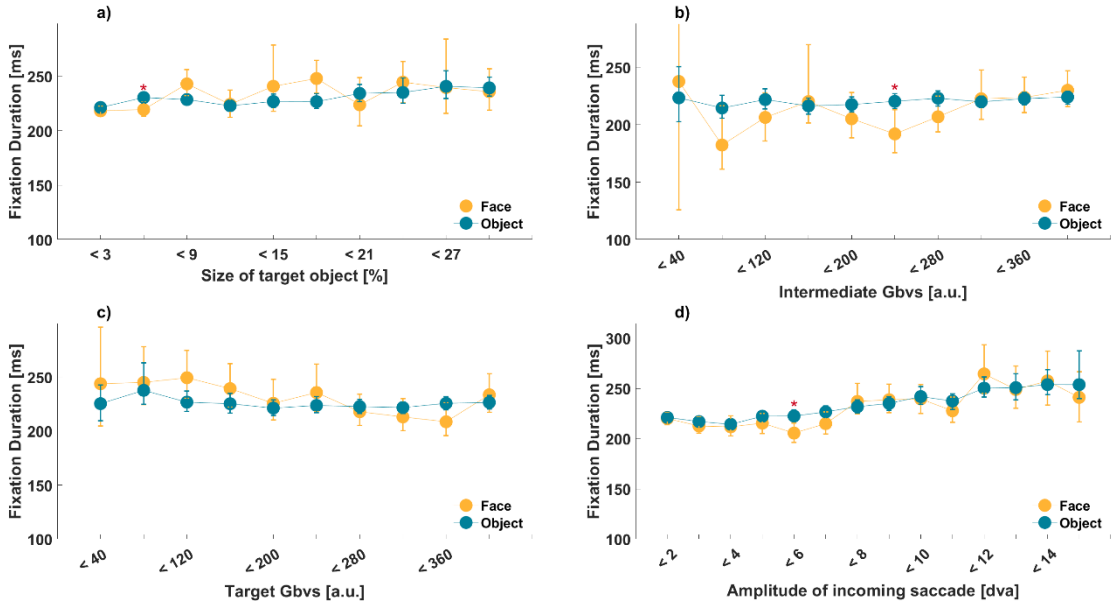

Figure S3. **Differences between intermediate fixation durations preceding face- vs. inanimate object-directed saccades.** Red asterisks mark Bonferroni corrected significance of paired t-test and error bars represent bootstrapped 95 % confidence

interval (1,000 resamples). Panel (a) shows fixation duration as a function of size of target stimuli (Size). Panel (b) shows fixation duration as a function of low-level salience at the intermediate fixation. Panel (c) shows fixation duration as a function of low-level salience at the target fixation. Panel (d) shows fixation duration as a function of amplitude of the incoming saccade. Low-level salience is given in arbitrary units (a.u.); low values indicate low saliency and high values indicate high saliency. Cyan and yellow markers denote data from inanimate object and face-directed saccades as shown in the inset.

4 Control Analysis: Amplitude

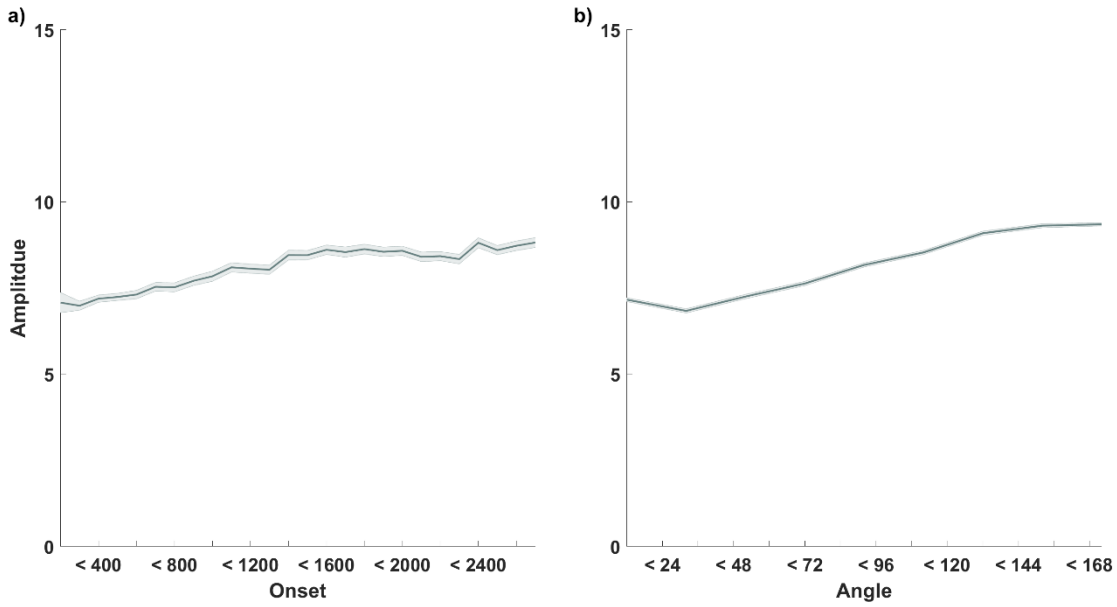

Figure S4. **Target amplitude as a function of trial time and relative saccadic angle**. Panel (a) shows the amplitude of target saccades as a function of the time across trial duration. Mean amplitudes were calculated using a sliding window with a width of 100 ms. Panel (b) shows target amplitude as a function of the angle between incoming and target saccade. Mean amplitudes were calculated using a sliding window with a width of 40 deg. Shaded areas represent the standard error of the mean.

Amplitude			
Fixed effects	Estimate	SE	t
(Intercept)	-0.010	0.013	-0.810 n.s.
Face	0.314	0.012	25.450*
Onset	0.084	0.004	19.540*
Angle	0.204	0.004	47.020*
Log-likelihood	-66396.400		
Deviance	132792.800		
AIC	132804.800		
BIC	132857.600		

** p<.05; n.s. = non-significant (p>.05)*

Table S9. Linear mixed-effects models fitting standardized amplitude. The table shows simple main effects of sematic category (Face), time from trial onset (Onset), and angle of target amplitude to incoming amplitude (Angle). Asterisks in the column of t-values indicate statistically significant beta coefficients. All continuous predictors were z-scored.

5 Target Size

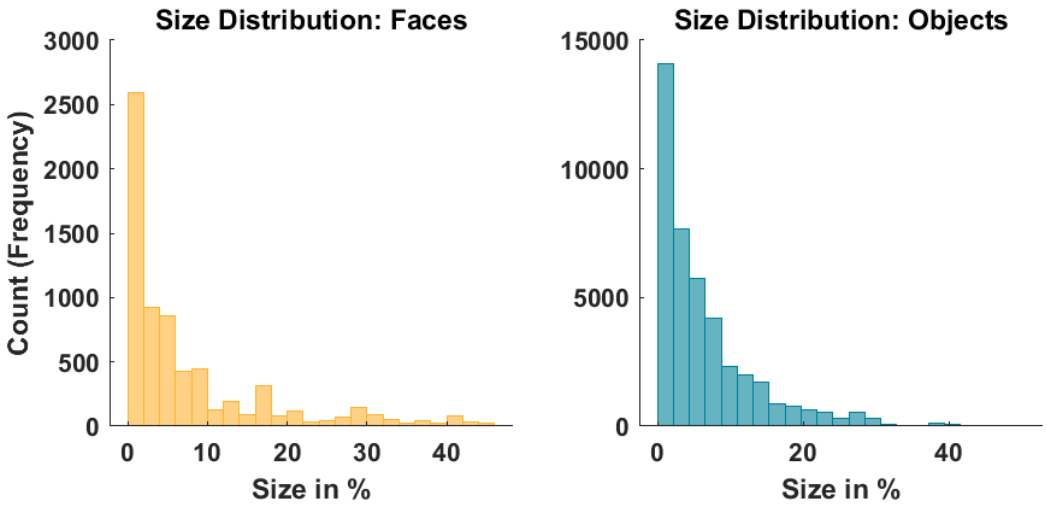

Figure S6. Distribution of the size of inner faces and inanimate objects. The figure shows the number of inner face masks and inanimate object masks by their size expressed as % of the total image size. Images size was 29.7 x 22.3 degrees visual angle and the majority of faces was not larger than 3 dva.

6 Animacy in Images

	Total	Nr images including
Faces	1204	469
People	1663	459
Animals	367	148

Total number of images = 700

Table S10. Frequency of faces, people, and animals in the images. The total number of images was 700, out of that 469 contained faces, 459 contained people, and 148 contained animals. No intermediate or target fixations falling on animals were included.

7.2 Study 2

Borovska, P., & de Haas, B. (2024). Individual gaze shapes diverging neural representations. *Proceedings of the National Academy of Sciences*, *121*(36), 2017. https://doi.org/10.1073/pnas.2405602121

PNAS BRIEF REPORT | PSYCHOLOGICAL AND COGNITIVE SCIENCES OPEN ACCESS

Individual gaze shapes diverging neural representations

Petra Borovska[a,1] and Benjamin de Haas[a,b]

Edited by Renée Baillargeon, University of Illinois at Urbana-Champaign, Champaign, IL; received March 21, 2024; accepted August 4, 2024

Complex visual stimuli evoke diverse patterns of gaze, but previous research suggests that their neural representations are shared across brains. Here, we used hyperalignment to compare visual responses between observers viewing identical stimuli. We find that individual eye movements enhance cortical visual responses but also lead to representa- tional divergence. Pairwise differences in the spatial distribution of gaze and in semantic salience predict pairwise representational divergence in V1 and inferior temporal cortex, respectively. This suggests that individual gaze sculpts individual visual worlds.

individual differences | gaze behavior | complex visual stimuli | hyperalignment | fMRI

Do individual brains represent the visual world in idiosyncratic ways? Is our view of complex visual stimuli unique? Given the foveal bias of the inferior temporal cortex (1, 2), ventral stream representations may be shaped by systematic idiosyncrasies of individual gaze (3) for static and dynamic stimuli (3–5). On the other hand, a range of previous results suggests that individual gaze may not matter much for the representation of com- plex, naturalistic stimuli: Eye-tracking studies have found higher interpersonal coherence for gaze toward directed content (6, 7) and stimuli with salient motion (8). Neuroimaging studies found that visually evoked brain activation in the ventral stream is only weakly modulated by eye movements (9) and may reflect broader visual field coverage (10) than previously assumed (1, 11, 12) and that individual representations of movie stimuli can be aligned across observers via linear transformations (13–16).

Here, we set out to test whether different individuals freely watching the same movie may nevertheless have individually divergent neural representations, which can be explained by idiosyncrasies in gaze. Specifically, we test whether the linear alignment of their neural representations is lower when observers freely watch a movie, as compared to fixating centrally while the movie is playing. Additionally, we test whether the degree of this representational divergence can be explained by systematic individual differences in gaze parameters for a given pair of observers.

Results

We let participants watch the same movie twice ($N = 19$, *SI Appendix*, *Supporting Methods* for sample details), once in an eye tracker and once in an fMRI session, the order of which was counterbalanced across participants. We tested effects of divergent gaze on neural representations, focusing on IT and V1 (Fig. 1*A*). Data from the scanning session were used for hyperalignment (15) to compare neural responses across observers: We fitted a linear transfer function to map the neural responses of one observer onto those of another and then applied it to holdout data of the same two participants and tested how well we could decode movie snippets. That is, we identified which movie snippet evoked a given response in an observer's brain, based on predictions generated from responses in the other observer's brain and the transfer function learned from independent data. Specifically, we used a Nearest Neighbor approach, identifying the highest correlating prediction across a library of predictions for hundreds of movie snippets. The accuracy of this cross-brain decoding was determined separately for each pair of observers and region of interest (ROI). It served as a proxy for the neural alignment of movie representations (*SI Appendix*, *Supporting Methods*).

To test the hypothesis that individual gaze leads to diverging neural representations, we tested two conditions in the fMRI session, instructing observers to either freely watch the movie or fixate centrally while it was playing. Cross-brain decoding dramatically decreased for free viewing compared to central fixation in V1 (24% vs. 45%, $b = -0.2$, $SE = 0.01$, $t(682) = -15.83$, $P < 0.001$; Fig. 1 *A*, *Bottom*) and IT (24% vs. 28%, $b = -0.06$, $SE = 0.009$, $t(682) = 6.96$, $P < 0.001$; Fig. 1 *A*, *Top*), as determined with a generalized linear mixed-effects model (*SI Appendix*, *Supporting Methods*). This shows that individual gaze leads to stronger but more idiosyncratic responses in V1 and IT, despite a significant amplitude increase of BOLD responses in V1 ($t(18) = 3.53$, $P < 0.01$), IT ($t(18) = 4.32$, $P < 0.001$), and beyond (Fig. 1*D*) for free-viewing compared to central fixation.

Author affiliations: [a]Department of Experimental Psychology, Justus Liebig University, Giessen 35394, Germany; and [b]Center for Mind, Brain and Behavior, Marburg and Giessen, Darmstadt 35032, Germany

Author contributions: B.d.H. designed research; P.B. performed research; P.B. analyzed data; B.d.H. administered and supervised the project; and P.B. and B.d.H. wrote the paper.

The authors declare no competing interest.

[1]To whom correspondence may be addressed. Email: Petra.Borovska@psychol.uni-giessen.de.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2405602121/-/DCSupplemental.

Published August 30, 2024.

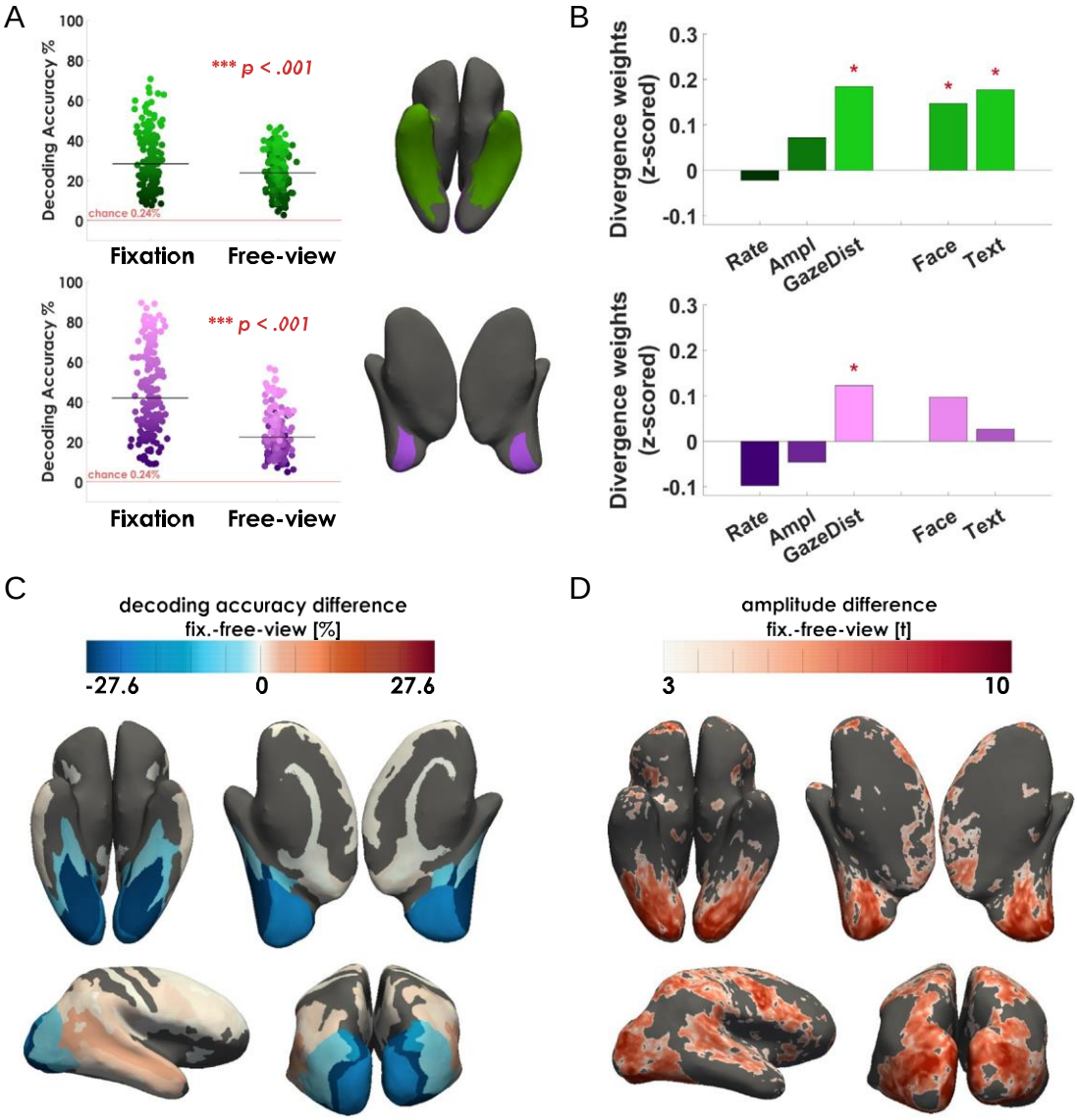

Fig. 1. Results. (*A*) Cross-brain decoding accuracy in the fixation and free-viewing conditions for IT and V1. Each dot represents one pair of observers ($N = 171$ pairs), the chance level of 0.24 % is indicated by red lines, and *P*-values correspond to a GLMEs testing the effects of the conditions on cross-brain decoding accuracy. The *Top* (green) and *Bottom* (purple) plots show data from IT and V1 and the corresponding ROIs. The shades of dots correspond to decoding accuracy in the fixation condition. (*B*) Fitted weights of individual differences in low- and high-level gaze parameters predicting pairwise neural divergence in IT and V1. Simple main effects are shown for IT in green (*Top*) and V1 in purple (*Bottom*). Rate and Ampl: individual difference in the saccadic rate and amplitude; GazeDist: average Euclidean distance between gaze positions of two observers; Face and Text: individual difference in the tendency to fixate faces and text, respectively. Asterisks indicate significant simple main effects that survived Bonferroni correction. (*C*) Effects of free-viewing on neural alignment between observers. The heatmap shows the average difference in cross-brain decoding accuracy (in %) between free-viewing and fixation conditions across all pairs of observers on the inflated cortical surface. The color-coding indicates the magnitude of the difference for each region of interest (ROI), extracted using the Destrieux atlas parcellation and the Benson retinotopy atlas for V1, V2, and V3. Only differences that were significant at $P < 0.001$ according to the GLME are shown (*SI Appendix*, *Supporting Methods*). Values are color-coded as shown in the inset bar, with cool colors indicating a drop of decoding accuracy in the free-viewing condition. (*D*) Amplitude effects of free-viewing. The heatmap shows vertex-wise t-values for the contrast between free-viewing and fixation on the inflated cortical surface of an example observer. *t* values are color-coded as shown in the inset bar. Inflated hemispheres are shown in inferior (*Top Left*), medial (*Top Right*), lateral (*Bottom Left*), and posterior views (*Bottom Right*).

To further analyze the ways in which gaze shapes individual neural representations, we calculated pairwise differences for a range of gaze parameters during the eye-tracking session. We then tested the contribution of these pairwise differences in gaze to the pairwise cross-brain decoding accuracy in the free viewing condition in IT and V1 (Fig. 1*B*), while controlling for the corresponding accuracy in the fixation condition. To illustrate the contribution of differences in gaze to representational divergence, we flipped the sign of best fitting weights in this linear regression such that positive weights indicate a contribution to lower cross-brain accuracy during free-viewing.

Among low-level predictors, the pairwise Euclidean distance of gaze positions in the eye-tracking session significantly predicted representational divergence in the scanning session in both, V1 ($b = 0.12$, $SE = 0.04$, $t(148) = 2.51$, $P < 0.05$; Fig. 1 *B*, *Bottom*) and IT ($b = 0.18$, $SE = 0.05$, $t(148) = 3.29$, $P < 0.01$; Fig. 1 *B*, *Top*), but pairwise differences in saccadic amplitude and rate did not (all $|b| \leq 0.09$, $|t| \leq 2.00$, $P \geq 0.09$; all *P*-values Bonferroni corrected, *SI Appendix*, *Supporting Methods*). Regarding semantic salience biases (3), pairwise differences in the tendency to fixate faces ($b = 0.14$, $SE = 0.04$, $t(149) = 2.99$, $P < 0.01$) and text ($b = 0.17$, $SE = 0.05$, $t(149) = 3.48$, $P < 0.01$; Fig. 1 *B*, *Top*) both significantly predicted pairwise representational divergence in IT, but not V1 (all $|b| \leq 0.09$, $|t| \leq 2.02$, $P \geq 0.08$, all *P*-values Bonferroni corrected, *SI Appendix*, *Supporting Methods*).

Taken together, individual gaze led to enhanced but also more diver- gent neural responses in the early visual cortex and IT. The pairwise

Euclidean distance of gaze positions during the eye-tracking session significantly predicted neural divergence in the later scanning session for both V1 and IT, whereas pairwise differences in the tendency to fixate faces and text significantly predicted neural divergence in IT only. To explore the effects of divergent gaze more comprehensively, we expanded the comparison of cross-brain decoding accuracy between free viewing and central fixation to anatomical parcels encompassing the entire brain (*SI Appendix*, *Supporting Methods*). Results showed that the effect of neural divergence was largely confined to occipital and inferior temporal regions, with stronger effects in more posterior parts of IT and early visual areas (Fig. 1*C*).

Discussion

Previous research has demonstrated that representations of complex visual stimuli in the human inferior temporal cortex (IT) can be decoded across brains using hyperalignment (13, 17). Here, we tested the hypothesis that such cross-brain decoding is limited by individual differences in gaze, leading to representational divergence.

Our findings show that free-viewing (compared to central fixation) leads to a strong increase of BOLD signal amplitudes across the visual system. This aligns with studies in humans and primates, showing enhanced neural responses to gaze shifts updating the visual input (7, 9, 18). Nevertheless, cross-brain decoding accuracy dramatically decreased in the free viewing condition, most prominently in the early visual cortex and posterior IT.

Moreover, pair-wise differences of gaze parameters in the eye-tracking session predicted the degree of gaze-induced representational divergence in the scanner. This was true for the average Euclidean distance between gaze positions in both V1 and IT, as well as for individual differences in the tendency to foveate faces and text in IT. This resonates with the foveal bias of face- and text-preferring neural populations in IT (1, 19, 20) and strongly suggests that individual gaze shapes individual visual worlds (3). Note that eye-tracking and fMRI sessions took place on different days and their order was counterbalanced across participants. Individual differences in gaze were highly consistent during the eye-tracking session, and previous studies have shown trait-like individual gaze biases (21, 22) which are shaped by genes (23, 24) and experience (25) and may be tuned to idiosyncratic characteristics of the visual system (26–28). Here, we find that individual gaze leads to systematic divergence of neural representations for identical complex visual stimuli. This matches recent results showing a relationship between indi- vidual differences in scene descriptions and gaze (29). Future studies should trace the developmental interplay between idio- syncrasies of the individual visual system, gaze, and perception. Furthermore, future research is needed to understand the con- sequences of representational divergence for communication, cooperation, individual preferences, and skills.

Materials and Methods

Participants watched "Shaun the Sheep" in two sessions, once in an eye tracker (N = 39; M_{age} = 23.25; SD = 3.87; 27 females) and a subset of participants in an fMRI scanner (N = 19; M_{age} = 24.68; SD = 3.68; 12 females), with the order of experiments counterbalanced. Eye-tracking data were collected using an Eyelink 1000 (SR Research, Ottawa, Canada). The fMRI session used a 3-Tesla Siemens Prisma. Functional images were preprocessed with realignment, coregistration, smoothing, temporal filtering, and further denoising (*SI Appendix*, *Supporting Methods*). All participants provided written informed consent, and the study was approved by the local ethics committee of Justus-Liebig-University Giessen and in accord with the Declaration of Helsinki.

Data, Materials, and Software Availability. Anonymized Eye-tracking and fMRI data have been deposited in Open Science Framework (https://osf.io/3pkw2/) (30).

ACKNOWLEDGMENTS. We would like to thank Dr. Arash Akbarinia for applying the DNN algorithm EAST for frame-wise labeling of text features in videos, Maximilian D. Broda for his help with implementation of the fMRI experiment, and Diana Weissleder, Jule Vorndamm, Lovis Rosenbaum, and Caroline Stephanie Wunn for their help with data collection. MR imaging for this study was performed at the Bender Institute of Neuroimaging at the Justus Liebig University Giessen, Germany. This work was supported by European Research Council Starting Grant 852885 INDIVISUAL and Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) Project No. 222641018-SFB/TRR 135 TP C9. BdH was further supported by "The Adaptive Mind," funded by the Excellence Program of the Hessian Ministry of Higher Education, Science, Research and Art

1. U. Hasson, I. Levy, M. Behrmann, T. Hendler, R. Malach, Eccentricity bias as an organizing principle for human high-order object areas. *Neuron* **34**, 479–490 (2002).
2. D. J. Kravitz, K. S. Saleem, C. I. Baker, L. G. Ungerleider, M. Mishkin, The ventral visual pathway: An expanded neural framework for the processing of object quality. *Trends Cogn. Sci.* **17**, 26–49 (2013).
3. B. de Haas, A. L. Iakovidis, D. S. Schwarzkopf, K. R. Gegenfurtner, Individual differences in visual salience vary along semantic dimensions. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 11687–11692 (2019).
4. M. D. Broda, B. de Haas, Individual fixation tendencies in person viewing generalize from images to videos. *Iperception* **13**, 1–10 (2022).
5. M. Rubo, M. Gamer, Social content and emotional valence modulate gaze fixations in dynamic scenes. *Sci. Rep.* **8**, 1–12 (2018).
6. U. Hasson *et al.*, Neurocinematics: The neuroscience of film. *Projections* **2**, 1–26 (2008).
7. K. H. Lu, S. C. Hung, H. Wen, L. Marussich, Z. Liu, Influences of high-level features, gaze, and scene transitions on the reliability of BOLD responses to natural movie stimuli. *PLoS One* **11**, 1–19 (2016).
8. M. Dorr, T. Martinetz, K. R. Gegenfurtner, E. Barth, Variability of eye movements when viewing dynamic natural scenes. *J. Vis.* **10**, 1–17 (2010).
9. S. Nishimoto, A. G. Huth, N. Y. Bilenko, J. L. Gallant, Eye movement-invariant representations in the human visual system. *J. Vis.* **17**, 1–10 (2017).
10. J. Park *et al.*, Immersive scene representation in human visual cortex with ultra-wide-angle neuroimaging. *Nat. Commun.* **15**, 1–15 (2024).
11. K. Grill-Spector, K. S. Weiner, The functional architecture of the ventral temporal cortex and its role in categorization. *Nat. Rev. Neurosci.* **15**, 536–548 (2014).
12. B. R. Conway, The organization and operation of inferior temporal cortex. *Annu. Rev. Vis. Sci.* **4**, 381–402 (2018).
13. J. V. Haxby, J. S. Guntupalli, S. A. Nastase, M. Feilong, Hyperalignment: Modeling shared information encoded in idiosyncratic cortical topographies. *Elife* **9**, 1–26 (2020).
14. M. Feilong, S. A. Nastase, J. S. Guntupalli, J. V. Haxby, Reliable individual differences in fine-grained cortical functional architecture. *Neuroimage* **183**, 375–386 (2018).
15. J. V. Haxby *et al.*, A common, high-dimensional model of the representational space in human ventral temporal cortex. *Neuron* **72**, 404–416 (2011).
16. M. Visconti, V. Chauhan, G. Jiahui, M. I. Gobbini, An fMRI dataset in response to "The Grand Budapest Hotel", a socially-rich, naturalistic movie. *Sci. Data* **7**, 1–9 (2020).
17. J. V. Haxby *et al.*, A common, high-dimensional model of the representational space in human ventral temporal cortex. *Neuron* **72**, 404–416 (2011).
18. W. Xiao, S. Sharma, G. Kreiman, M. S. Livingstone, Feature-selective responses in macaque visual cortex follow eye movements during natural vision. *Nat. Neurosci.* **27**, 1157–1166 (2024), 10.1038/s41593-024-01631-5.
19. K. Grill-Spector, K. S. Weiner, K. Kay, J. Gomez, The functional neuroanatomy of human face perception. *Annu. Rev. Vis. Sci.* **3**, 167–196 (2017).
20. E. H. Silson, I. I. A. Groen, C. I. Baker, Direct comparison of contralateral bias and face/scene selectivity in human occipitotemporal cortex. *Brain Struct. Funct.* **227**, 1405–1421 (2022).
21. B. de Haas, A. L. Iakovidis, D. S. Schwarzkopf, K. R. Gegenfurtner, Individual differences in visual salience vary along semantic dimensions. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 11687–11692 (2019).
22. M. D. Broda, B. de Haas, Individual differences in looking at persons in scenes. *J. Vis.* **22**, 1–12 (2022).
23. J. N. Constantino *et al.*, Infant viewing of social scenes is under genetic control and is atypical in autism. *Nature* **547**, 340–344 (2017).
24. D. P. Kennedy *et al.*, Genetic influence on eye movements to complex scenes at short timescales. *Curr. Biol.* **27**, 3554–3560.e3 (2017).
25. M. Linka, Ö. Sensoy, H. Karimpur, G. Schwarzer, B. de Haas, Free viewing biases for complex scenes in preschoolers and adults. *Sci. Rep.* **13**, 1–15 (2023).
26. M. F. Peterson, M. P. Eckstein, Individual differences in eye movements during face identification reflect observer-specific optimal points of fixation. *Psychol. Sci.* **24**, 1216–1225 (2013).
27. M. D. Broda, B. de Haas, Individual differences in human gaze behavior generalize from faces to objects. *Proc. Natl. Acad. Sci. U.S.A.* **121**, e2322149121 (2023).
28. M. J. Arcaro, P. F. Schade, J. L. Vincent, C. R. Ponce, M. S. Livingstone, Seeing faces is necessary for face-domain formation. *Nat. Neurosci.* **20**, 1404–1412 (2017).
29. D. Kollenda, A. V. Reher, B. de Haas, Individual gaze predicts individual scene descriptions PsyArxiv [Preprint]. (2024). https://doi.org/10.31234/osf.io/nx7jy (Accessed 15 May 2024).
30. P. Borovska, B. de Haas, Individual Gaze Shapes Diverging Neural Representations. OSF. https://osf.io/3pkw2. Deposited 18 June 2024.

7.2.1 Supplement: Study 2

Supporting Methods

Participants

A total of 42 healthy participants with normal or corrected-to-normal vision took part in the eye-tracking experiment. The datasets of 3 participants were excluded (one due to excessive head movement, two because of missing datafiles), leaving a sample of 39 participants (M_{age} = 23.25; *SD* = 3.87; 27 females). In a separate fMRI experiment, we recorded brain activity of 22 of these subjects. The dataset of 3 participants were excluded due to technical difficulties in the scanner, leaving a sample of 19 participants (M_{age} = 24.68; *SD* = 3.68; 12 females). One additional participant was excluded from the analysis investigating the effect of gaze parameters on representational divergence, because of a recording failure in the eye-tracking session. All participants provided written informed consent, the study was approved by the local ethics committee of Justus-Liebig-University Giessen and in accord with the Declaration of Helsinki. Participants could choose between course credit or 10€/h for their participation.

Design

Participants watched the audio-visual feature movie Shaun the Sheep in two separate sessions. The final sample of 19 participants saw the movie twice (on different days), once in the scanner and once in a separate session in an eye-tracker. The order of sessions was counter-balanced across participants (with 11 included participants in the eye-tracker first). An additional 20 participants watched the movie in the eye-tracker only.

Eye-tracking experiment

Apparatus

Participants sat in a dark room with their head in a chin- and forehead-rest at a distance 55 cm from the screen. Gaze data were acquired using a tower mount Eyelink 1000 (SR Research, Ottawa, Canada) at a frequency of 1kHz. Stimuli were shown on a 23.8-inch LG Ultra HD monitor at a resolution of 3840 × 2160 pixels and a refresh rate of 59 Hz. Participants viewed the stimuli at a distance of ~55 cm and size of 48 x 27 degrees visual angle (dva). The experiment

was programmed using Psychtoolbox version 3.0.16 (Kleiner et al., 2007) in MATLAB R2021b (Mathworks, Natick, MA, USA) on a Windows 10 PC.

Procedure

Participants watched in the eye-tracker the first 50 minutes of the movie, divided into 10 blocks of 5 minutes each. Each block started with a calibration. Participants were instructed to freely view the movie, keep their head still and pay attention.

Functional MRI experiment

Procedure

The scanning session was divided into three runs, each lasting approximately ~21 mins. Each run consisted of four blocks presenting 5 mins of the movie. Participants were instructed at the beginning of each block to either fixate a super-imposed dot at the center of the movie or free-view for the following 5 mins. Each run had the same order of condition blocks: fixation, free-viewing, fixation, free-viewing. This sequence repeated for each run. At the end of each block, a blank screen was shown for 20s. In addition to these functional scans, we recorded an anatomical scan and field map for each participant.

Data acquisition

The MRI session was carried out on a 3-Tesla imaging system (Siemens Prisma) with a 64-channel head coil at the Bender Institute of Neuroimaging (BION) at Giessen University. Stimuli were shown on a projector Epson EB-G5600 with a resolution of 1024 x 768 pixels and a refresh rate of 60 Hz. The stimulus video was rescaled to approximate the size of the stimulus in the Eyetracker with a width of ~48 dva and a height of ~27 dva. Eye movements were monitored online using a ViewPoint Eye-camera (Arrington Research Inc., Scottsdale, AZ) to ensure that participants followed the instructions in both conditions (fixating vs. free-viewing the movie).

Functional images covered the whole brain and were obtained using a multiband echo-planar imaging sequence (EPI) with an echo time (TE) of 33 ms and a repetition time (TR) of 1000 ms. Further parameters for obtaining functional data were as follows: Field of view (FoV) = 240 × 240 mm, in-plane resolution = 2.5 mm × 2.5 mm, 52 sagittal slices (descending) with a thickness

of 2.5 mm and a distance factor of 20%, flip angle (FA) = 59°, acceleration factor = 4. Per participant, 3990 volumes of functional data were acquired (1330 per run).

High-resolution anatomical images were obtained using a T1-weighted magnetization-prepared rapid acquisition gradient-echo (MPRAGE) sequence with the following scan parameters: FoV = 240 × 240 mm, TE = 3.53 ms, TR = 1880 ms, inversion time = 949 ms, in-plane resolution = 0.94 mm × 0.94 mm, number of slices = 176, slice thickness = 0.94 mm, flip angle (FA) = 8°.

Magnetic field perturbations were accounted for by measuring a field map with the following scan parameters: FoV = 220 × 220 mm, TE (1) = 10 ms, TE (2) = 12.46 ms, TR = 1,000 ms, in-plane resolution = 2.0 × 2.0 mm, slice thickness = 3.0 mm, number of slices = 40 (transversal), FA = 90°.

Preprocessing

All image files were converted to NIfTI format and preprocessed using SPM 12 (http://www.fil.ion.ucl.ac.uk/spm/software/spm12/) and custom MATLAB code. The remaining functional images were realigned and unwarped using the voxel displacement maps generated from the field maps. Further, the functional images were co-registered to the structural scan and spatially smoothed using a 4 mm Gaussian kernel (Guntupalli et al., 2016). Time series were bandpass filtered with a discrete cosine transform removing slow drifts and high frequency noise with a cut-off of 1/128 Hz, and subsequently z-scored. To remove noise, we extracted six rigid-body motion parameters and framewise displacements (all estimated during realignment), as well as three principal components from cerebrospinal fluid and white matter time-series. These nuisance regressors were regressed out from functional data for each run (Visconti di Oleggio Castello et al., 2020). Finally, the first six volumes of all functional images were discarded to account for delays in the hemodynamic response.

Data analysis

Gaze consistency measures

To quantify gaze parameters, we extracted saccades and fixations using the SR Research saccade detection algorithm (velocity >30 degrees/s and acceleration >8000 degrees/s2). Gaze coordinates that fell outside of video borders were excluded. Additionally, fixations with a

duration under 100 ms were excluded (SR Research, 2022). This led to an exclusion of less than 1 % of fixations on average. To prevent erroneous gaze estimation during lid occlusion caused by a blink, saccades occurring 100 ms before or after a blink were also discarded (i.e., ~7 % of saccades on average were removed). Additionally, saccade and fixations with a duration > 1,000 or peak velocity > 1000 deg/s were removed.

We used several publicly available DNN algorithms to label the movie stimuli on a frame-wise basis. To label text, we used EAST (An Efficient and Accurate Scene Text Detector (Zhou et al., 2017)). For face labels, we used YOLO5Face (Qi et al., 2023), and finally to label all remaining objects, we used YOLOV5 (Jocher, 2020). We discarded all object-labels that were overlapping with text and face.

We included several gaze parameters for this and subsequent parts of the analysis. Fixations that fell within a distance of 0.5 dva from a given label, were marked accordingly, as face, text or other object fixations. We then calculated the proportion of text and face fixations among all labelled fixations for each observer. Similarly, we computed the median saccadic rate and amplitude for each observer. To assess the consistency of individual differences, we correlated individual gaze parameters across odd and even splits of the movie. Furthermore, gaze parameters served as basis for calculating pairwise differences used in linear regression models (see below, Effect of gaze parameters on cross-decoding accuracy).

We established that individual gaze varies in highly systematic ways, even for a directed movie (Shaun the Sheep). Individual differences in low-level parameters of viewing dynamics were large (variability up to factor 4) and highly consistent (split half consistency of saccadic rate $r(39) = .97$, $p < .001$; saccadic amplitude $r(39) = .94$, $p < .001$). Similarly, individual fixation biases towards faces and text varied up to factor 2 and showed moderate to good consistency (face $r(39) = .76$, $p < .001$; text $r(39) = .32$, $p < .05$).

Response amplitudes

We specified boxcar regressors for blocks of fixation and free-viewing for each of the three runs and convolved them with the canonical hemodynamic response function, as implemented in SPM12. Each block had a duration of 5 mins. Additionally, we incorporated six motion parameters estimated during realignment. The conditions were contrasted against each other in a

general linear model (free-view > fixation), generating *t*-maps for each participant (i.e. first level analysis). These *t*-maps were subsequently masked with the IT region of interest (ROI). ROI masks were individually defined for each hemisphere using the FreeSurfer (http://surfer.nmr.mgh.harvard.edu) parcellation algorithm (Destrieux et al., 2010) and included the following labels: G_oc-temp_lat-fusifor, G_oc-temp_med-Parahip, G_temporal_inf, Pole_temporal, S_collat_transv_ant, S_oc-temp_lat, S_oc-temp_med&Lingual, and S_temporal_inf. Finally, the masked *t*-maps were averaged for each participant and tested against zero using a one-sample *t*-test.

Hyperalignment and cross-decoding accuracy

For cross-brain decoding, we used a modified version of the hyperalignment technique (Haxby et al., 2011b; Haxby, Guntupalli, et al., 2020). Unlike classical hyperalignment, we directly fitted the data from one participant's cortical anatomy onto that of another, without projecting them into a common space. Hyperalignment uses Procrustes transformations to align the voxel spaces of individual subjects to one another based on their responses during movie watching (Haxby et al., 2011). We ran the classification algorithm separately for each condition and pair of observers. We calculated the cross-brain decoding accuracy in an a priori defined IT mask combining the labels listed above (see Response amplitudes) and in a mask for V1 (based on Benson retinotopy atlas). We used a similar approach to calculate cross-brain decoding accuracy in the whole brain by running our pipeline for each ROI separately. ROIs were extracted using the Freesurfer parcellation according to the Destrieux atlas and combined with the Benson retinotopy atlas for V1, V2, and V3 (Avesani et al., 2019; Benson et al., 2012, 2014). This resulted in a total of 78 ROIs.

(1) In the first step (Learn) one member of the pair was assigned source and the other target. We divided the data for each condition into training and test sets of equal size. For voxel selection, we determined the highest correlation of each voxel in each ROI across its correlations with all voxels in the other brain's ROI. We then ranked voxels in each ROI based on this correlation score and selected the top 1000 voxels in both the target and source ROI for further analysis steps. Note that correlation scores and voxel selection were purely based on the training portion of data, but applied to the test data as well (cf. ref. Haxby et al., 2011). Then, we used Procrustes transformation to find the best fitting linear transfer function from the training data of the source

subject onto that of the target subject. (2) In the second step (Predict), we applied the resulting transformation matrix to the test data of the source subject to predict the test responses of the target subject. (3) In the last step (Evaluate), we divided the test data of the target subject and the corresponding predictions into 422 snippets of 18s each. Time snippets were derived using a sliding window with a step size of 2 s, and therefore partially overlapping. In each iteration the prediction snippets that overlapped with the target test snippet were discarded resulting in total of 422 test snippets in each iteration (cf. ref. Haxby et al., 2011). Then, we determined the Nearest Neighbor prediction for each target test snippet as the prediction snippet correlating the highest with the target test snippet. If the Nearest Neighbor prediction corresponded to the same time-window of movie watching this was registered as a decoding success, otherwise as a failure. Decoding accuracy corresponds to the percentage of successes. Across both folds of training and test data chance level corresponds to 0.24% (1/422 snippets). This procedure was repeated for all possible pairings of observers. Within each pair, decoding accuracy was averaged across both folds of training and test data and iterations in which either observer served as the target.

We performed a similar analysis for pre-defined IT and V1 ROIs, and extended our analysis to include all remaining ROIs extracted from the Destrieux and Benson atlases. Next, we tested for significant differences in cross-brain decoding accuracy between free-view and fixation conditions. Given the non-independence of the cross-brain decoding accuracy pairs (N = 171), we built Generalized linear mixed-effects models for each ROI to test an effect of conditions on cross-brain decoding accuracy. We included a categorical predictor indicating condition with two levels (Fixation and Free-view) and two random factors (Source and Target) expressing the identity of each subject, in which either observer served as the target. We confirmed that the cross-brain decoding accuracy significantly decreased for the free-view condition for several ROIs, while taking into account the identity of each subject. Additionally, we performed a conservative control analysis which reduced the degrees of freedom to the number of participants. For each observer, we averaged all pairwise instances of cross-brain decoding with this target observer. This is an index of how well this observer's activation patterns can be decoded on average, based on hyperalignment with all other brains. We did this separately for the free-viewing and central fixation conditions and entered the differences between conditions for each of our 19 observers into a one-sample *t*-test against zero, thus reducing degrees of

freedom to 18. This conservative approach confirmed highly significant effects for both IT ($t(18) = 5.37$, $p < .001$) and V1 ($t(18) = 10.14$, $p < .001$).

Here, we list results of the first 12 ROIs with the largest decrease in cross-brain decoding accuracy in the free-view condition (which were all in early visual cortex and IT): V2 (51 % vs. 25 %, $b = -0.26$, $SE = 0.01$, $t(682) = -19.35$, $p < .001$); V3 (51 % vs. 24 %, $b = -0.26$, $SE = 0.01$, $t(682) = -20.24$, $p < .001$); V1 (45 % vs. 24 %, $b = -0.2$, $SE = 0.01$, $t(682) = -15.83$, $p < .001$); Ligual gyrus (50 % vs. 23 %, $b = -0.26$, $SE = 0.01$, $t(682) = -21.13$, $p < .001$); Occipital pole (37 % vs. 9 %, $b = -0.27$, $SE = 0.01$, $t(682) = -26.26$, $p < .001$); Superior occipital gyrus (34 % vs. 18 %, $b = -0.16$, $SE = 0.01$, $t(682) = -15.31$, $p < .001$); Middle occipital gyrus (27 % vs. 22 %, $b = -0.04$, $SE = 0.01$, $t(682) = -4.95$, $p < .001$); Posterior transverse collateral sulcus (25 % vs. 7 %, $b = -0.17$, $SE = 0.007$, $t(682) = -22.64$, $p < .001$); Calcarine sulcus (23 % vs. 16 %, $b = -0.07$, $SE = 0.007$, $t(682) = -9.2$, $p < .001$); Cuneus (21 % vs. 12 %, $b = -0.08$, $SE = 0.007$, $t(682) = -10.51$, $p < .001$); Lateral occipito-temporal gyrus (20 % vs. 13 %, $b = -0.06$, $SE = 0.007$, $t(682) = -8.57$, $p < .001$); and Inferior occipital gyrus and sulcus (19 % vs. 12 %, $b = -0.07$, $SE = 0.007$, $t(682) = -10.33$, $p < .001$). All significant differences between conditions for all ROIs are displayed in the main figure (see Figure 1c).

To estimate the success of hyperalignment, we additionally conducted correlation based Nearest Neighbor classification on normalized data in IT, without using hyperalignment. First, we registered the functional data of each subject into standard MNI space using SPM12. Then, we repeated our classification pipeline (specified above) leaving out the Procrustes transformation to align brains of the pair of observers. This procedure resulted in a drastic drop of cross-brain decoding accuracy in both conditions (2.1 % in the fixation condition; 1.7 % in the free-view condition). We tested for the statistical significance of an effect of condition on the cross-decoding accuracy using GLME. This revealed a significant decrease in cross-brain decoding accuracy in the free-view compared to the fixation condition ($b = -0.003$, $SE = 0.0009$, $t(682) = -3.84$, $p < .001$).

Effect of gaze parameters on cross-decoding accuracy

To assess the impact of gaze parameters on representational divergence, we specified separate multiple linear regression models for IT and V1. Both models tested pairwise differences in gaze parameters during the eye-tracking session as predictors of pairwise representational divergence

during the scanning session. Specifically, we regressed pairwise differences in gaze parameters onto pairwise cross-brain decoding accuracy in the free-viewing condition, controlling for the cross-brain decoding accuracy in the fixation condition. Then we flipped the sign of the resulting best fitting weights to display the contribution of diverging gaze to representational divergence (i.e. positive weights indicating a decrease in cross-brain decoding in the free-viewing condition). The first, low-level model included unsigned inter-observer differences in median saccadic amplitude and rate, as well as the median Euclidean distance of gaze positions as predictors. The second, high-level model, included unsigned individual differences in the proportion of labelled fixations falling onto faces and text. All predictors of interest except the Euclidean distance of gaze positions were normalized by their average across a given pair (to express differences relative to the overall magnitude of a given trait; for instance, the difference in saccadic rate between two observers was normalized by the average of the saccadic rate across these two observers). This was repeated for all pairs of observers, resulting in observer *dissimilarity* matrices for each gaze parameter and a corresponding matrix of pairwise decoding accuracies, each of which was vectorised and z-scored before being entered into the model. Resulting *p*-values were Bonferroni corrected for three predictors in the low-level model and two predictors in the high-level model. models (two for each ROI).

Controlling for decoding accuracy in the fixation condition resulted in a significant effect on neural divergence in free-view condition for all models: low-level model in IT ($b = 0.72$, $SE = 0.05$, $(148) = 13.002$, $p < .001$); high-level model in IT ($b = 0.72$, $SE = 0.05$, $t(149) = 14.32$, $p < .001$); low-level model in V1 ($b = 0.78$, $SE = 0.04$, $t(148) = 16.34$, $p < .001$); high-level model in V1 ($b = 0.79$, $SE = 0.04$, $t(149) = 16.35$, $p < .001$).

1. Visconti di Oleggio Castello, M., Chauhan, V., Jiahui, G. & Gobbini, M. I. An fMRI dataset in response to “The Grand Budapest Hotel”, a socially-rich, naturalistic movie. *Sci Data* **7**, 1–9 (2020).
2. Zhou, X. *et al.* EAST: An efficient and accurate scene text detector. *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017* **2017-Janua**, 2642–2651 (2017).
3. Qi, D., Tan, W., Yao, Q. & Liu, J. YOLO5Face: Why Reinventing a Face Detector. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* **13805 LNCS**, 228–244 (2023).
4. Jocher, G. Ultralytics YOLOv5 (7.0). Preprint at https://doi.org/https://doi.org/10.5281/zenodo.3908559 (2020).
5. Destrieux, C., Fischl, B., Dale, A. & Halgren, E. Automatic parcellation of human cortical gyri and sulci using standard anatomical nomenclature. *Neuroimage* **53**, 1–15 (2010).
6. Haxby, J. V. *et al.* A common, high-dimensional model of the representational space in human ventral temporal cortex. *Neuron* **72**, 404–416 (2011).
7. Haxby, J. V., Guntupalli, J. S., Nastase, S. A. & Feilong, M. Hyperalignment: Modeling shared information encoded in idiosyncratic cortical topographies. *Elife* **9**, 1–26 (2020).
8. Benson, N. C., Butt, O. H., Brainard, D. H. & Aguirre, G. K. Correction of Distortion in Flattened Representations of the Cortical Surface Allows Prediction of V1-V3 Functional Organization from Anatomy. *PLoS Comput Biol* **10**, (2014).
9. Benson, N. C. *et al.* The retinotopic organization of striate cortex is well predicted by surface topology. *Current Biology* **22**, 2081–2085 (2012).
10. Avesani, P. *et al.* The open diffusion data derivatives, brain data upcycling via integrated publishing of derivatives and reproducible open cloud services. *Sci Data* **6**, 1–13 (2019).

8 List of all publications

Borovska, P., & de Haas, B. (2024). Individual gaze shapes diverging neural representations. Proceedings of the National Academy of Sciences, 121(36), 2017. https://doi.org/10.1073/pnas.2405602121

Borovska, P., & de Haas, B. (2023). Faces in scenes attract rapid saccades. Journal of Vision, 23(8), 1–15. https://doi.org/10.1167/jov.23.8.11

Broda, M. D., Borovska, P., Kollenda, D., Linka, M., de Haas, N., de Haas, S., & de Haas, B. (2024). Estimating the human bottleneck for contact tracing. PNAS nexus, 3(7), pgae283. https://doi.org/10.1093/pnasnexus/pgae283

Broda, M. D., Borovska, P., & de Haas, B. (2024). Individual differences in face salience and rapid face saccades. Journal of Vision, 24(6), 16. https://doi.org/10.1167/jov.24.6.16

9 Selbständigkeitserklärung

Hiermit erkläre ich, dass ich die vorliegende Arbeit selbständig und ohne unzulässige Hilfe oder Benutzung anderer als der angegebenen Hilfsmitel angefertigt habe. Alle Textstellen, die wörtlich oder sinngemäß aus veröffentlichten oder nichtveröffentlichten Schriften entnommen sind, und alle Angaben, die auf mündlichen Auskünften beruhen, sind als solche kenntlich gemacht. Bei den von mir durchgeführten und in der Dissertation erwähnten Untersuchungen habe ich die Grundsätze guter wissenschaftlicher Praxis, wie sie in der „Satzung der Justus-Liebig-Universität Gießen zur Sicherung guter wissenschaftlicher Praxis“ niedergelegt sind, eingehalten sowie ethische, datenschutzrechtliche und tierschutzrechtliche Grundsätze befolgt. Ich versichere, dass Dritte von mir weder unmittelbar noch mittelbar geldwerte Leistungen für Arbeiten erhalten haben, die im Zusammenhang mit dem Inhalt der vorgelegten Dissertation stehen, und dass die vorgelegte Arbeit weder im Inland noch im Ausland in gleicher oder ähnlicher Form einer anderen Prüfungsbehörde zum Zweck einer Promotion oder eines anderen Prüfungsverfahrens vorgelegt wurde. Alles aus anderen Quellen und von anderen Personen übernommene Material, das in der Arbeit verwendet wurde, oder auf das direkt Bezug genommen wird, wurde als solches kenntlich gemacht. Insbesondere wurden alle Personen genannt, die direkt und indirekt an der Entstehung der vorliegenden Arbeit beteiligt waren. Mit der Überprüfung meiner Arbeit durch eine Plagiatserkennungssoftware bzw. ein internetbasiertes Softwareprogramm erkläre ich mich einverstanden.

_____ _____

Datum Unterschrift